

## Il diritto verso “il mondo nuovo”. Le sfide dell’Intelligenza Artificiale

Antonio D’Aloia\*

THE LAW TO “THE NEW WORLD”. THE CHALLENGES OF ARTIFICIAL INTELLIGENCE

ABSTRACT: AI is much more than a complex challenge to the system of law. This technology has the potential to change the world and to ask ourselves on the ethical and anthropological questions concerning the deepest meaning of human experience. The contribution investigates the legal implications of the AI systems’ development, from the responsibilities related to these new forms of behavior, to the risks linked to the profiling and micro targeting of people, to the application of algorithmic models in judicial decisions, up to the most radical question: will it be necessary or simply possible foreseeing a legal personhood for the artificial “agents”?

KEYWORDS: artificial intelligence; legal personhood; responsibility; privacy; algorithmic justice

SOMMARIO: 1. Tra fiction e diritto. Alla ricerca di principi e regole per l’AI – 2. AI, Robotics, mondo “digitale” – 3. Perché il diritto deve occuparsi di AI. Dalla cibernetica giuridica al diritto per i sistemi che imparano da soli – 4. Segue: Macchine (sempre più) autonome e responsabilità – 5. AI, diritto, diritti. Opportunità e problemi – 6. Algoritmi e (rischi di) discriminazioni – 7. AI e processo. Quale spazio per la decisione algoritmica? – 8. Privacy e altre garanzie nel trattamento algoritmico dei dati – 9. AI, algoritmi e “società dell’anticipazione”. Problemi del processo democratico – 10. L’interrogativo “intrattabile”. Ci sarà una soggettività per gli AI systems o agents? – 11. Prime (e provvisorie ...) conclusioni. Il diritto di fronte ad una sfida radicale e non temporanea.

### 1. Tra fiction e diritto. Alla ricerca di principi e regole per l’AI

I contributi raccolti in questo fascicolo di Biolaw sono una parte delle relazioni e degli interventi svolti nell’ambito del Convegno su *Intelligenza artificiale e diritto. Come regolare un mondo nuovo* (Parma, 12 ottobre 2018).

Sono stati selezionati in rapporto al loro maggior grado di connessione e di incrocio con l’orizzonte tematico della bioetica e del biodiritto. Le questioni poste dall’AI costituiscono, almeno parzialmente, una delle frontiere nuove della bioetica, portano quasi a ripensarne l’appropriatezza della dimensione settoriale rispetto ad un contenitore più largo, che potrebbe essere descritto sinteticamente come il luogo plurale del confronto tra sviluppi scientifici e tecnologici e scienze sociali (il diritto tra queste)<sup>1</sup>.

\* Professore ordinario di Diritto Costituzionale e Biodiritto, Università degli Studi di Parma. Mail: [antonio.daloia@unipr.it](mailto:antonio.daloia@unipr.it). Contributo su invito.

<sup>1</sup> Su “AI Ethics as sub-field of applied ethics and technology”, cfr. *Draft Ethics Guidelines for Trustworthy AI*, predisposto dall’High Level Expert Group on Artificial Intelligence della Commissione Europea (d’ora in avanti

Al di là adesso di discussioni nominalistiche o di “regolamenti di confine”, il tema del rapporto tra AI e sistemi sociali e giuridici, fino a pochi anni fa limitato ad una sfera ristretta di specialisti, talvolta considerati anche un po' visionari, oggi sembra esplodere in tutta la sua complessità di implicazioni e di sfide “adattive”.

Il tema dell'AI ha vissuto di fiammate e di delusioni (si è parlato di “primavere” e “inverni”)<sup>2</sup>. Dentro questi cicli molto oscillanti, è emerso ad ogni modo l'impatto su tutta una serie di aspetti della vita sociale ed economica, e hanno cominciato a prendere corpo interrogativi a volte capaci di scavare nel profondo dell'esperienza umana e di mettere in gioco basilari assunti etici ed antropologici.

E questo non poteva non chiamare in causa il diritto, la sua ricerca, spesso faticosa, di dare un senso alla realtà, allo scorrere delle vicende umane.

«Come regolare un mondo nuovo» non è allora solo un omaggio alla fantasia distopica di Aldous Huxley<sup>3</sup>. La sensazione è che effettivamente l'AI sta cambiando (o comunque appare capace di cambiare) radicalmente il mondo sotto i nostri occhi. Quello che fino a poco tempo fa sembrava dominio quasi esclusivo della *science fiction* (da Blade Runner a 2001 - Odissea nello spazio, fino ai più recenti Matrix, Transcendence, Her, Ex machina, The Circle, e le serie Westworld e Black Mirror<sup>4</sup>), sta assumendo forme concrete, quelle delle *driverless cars*, di sistemi diagnostici e/o chirurgici sempre più accurati e precisi (con livelli differenti di supervisione umana e “ad autonomia crescente”<sup>5</sup>), di algoritmi in grado di fare il *risk assessment* finalizzato a decidere se un soggetto deve essere mantenuto in stato di custodia cautelare, oppure può essere messo in libertà in attesa di accertare la sua responsabilità in un processo, o di fare lo *scoring* per la valutazione di ammissibilità di un contratto assicurativo o di mutuo, e altre cose ancora (nell'ambito ad esempio della gestione delle infrastrutture di mobilità o delle politiche ambientali).

Invero, l'immaginario scientifico e letterario su queste possibilità tecnologiche è molto più risalente nel tempo rispetto agli esempi cinematografici prima richiamati. Basti pensare al Golem della leg-

---

DEG), del 18 dicembre 2018, 2 e 6, non a caso citando tra i documenti di riferimento di questo nuovo settore la Convenzione di Oviedo del 1997.

<sup>2</sup> Per una breve storia delle ricerche sull'AI, v. J. BERNSTEIN, *Uomini e macchine intelligenti*, Milano, 1990, 19 ss., 54 ss., che ricorda la Conferenza estiva del Dartmouth College del 1956, e il Gruppo di Intelligenza Artificiale del MIT (Massachusetts Institute of Technology) creato nel 1958 da J. McCarthy e M. Minsky (quest'ultimo fu anche il consulente di Stanley Kubrick nella lavorazione di 2001-Odissea nello Spazio). A proposito delle delusioni, sul Teorema di Tesler, secondo cui «l'IA è tutto ciò che non è ancora stato fatto», v. D. R. HOFSTADTER, *Godel, Escher, Bach: un'eterna ghirlanda brillante*, trad. it., Milano, 1984, 649. Secondo C. ACCOTO, *Il mondo dato*, Milano, 2017, 9, invece, «la stagione dell'inverno dell'intelligenza artificiale potrebbe essere ormai definitivamente alle nostre spalle. Il lungo periodo di disaffezione fatto anche di fallimenti clamorosi, di carenza o scarsità di fondi e di interesse per la ricerca e lo sviluppo scientifico e industriale dell'intelligenza artificiale sembra si stia chiudendo».

<sup>3</sup> Il riferimento è al celebre romanzo di A. HUXLEY, *Il Mondo nuovo* (1932). Di un nuovo mondo, a proposito dell'AI, parla L. ALEXANDRE, *La guerra delle intelligenze*, trad. it., Torino, 2018, 26.

<sup>4</sup> Cfr., per una rassegna sintetica di questi titoli, S. LENZI, G. CIOFALO, WIENER, *Goldrake e i Social Media. L'eredità della cibernetica tra scienza, tecnologia e comunicazione*, introduzione a N. WIENER, *La Cibernetica* (1948), Roma, 2017, 23-24.

<sup>5</sup> Vedi infra, in questa Rivista, il saggio di D. AMOROSO E G. TAMBURRINI, *I sistemi robotici ad autonomia crescente tra etica e diritto: quale ruolo per il controllo umano?*

genda praghese, o alla concezione cartesiana degli animali come “automi”<sup>6</sup>, e prima ancora ai racconti su automi nell’antico Egitto, in Cina, nella Grecia antica, all’Homunculus di Paracelso, alla testa parlante dei Templari, al “moderno Prometeo” di Mary Shelley, al leone “robotico” che Leonardo di Vinci regalò al re di Francia nel 1515, al turco che giocava a scacchi costruito da Wolfgang von Kempelen nella seconda metà del XVIII secolo<sup>7</sup>, al dibattito settecentesco (La Mettrie, D’Holbach, Helvétius) sul rapporto uomo-animale-macchina (accuratamente ricostruito da A. Punzi<sup>8</sup>) all’anatra di de Vaucanson<sup>9</sup>.

Cinema, letteratura, diritto. Più che una sequenza tra sistemi separati, sembra essere, sull’AI, una rincorsa a capire e a mettere a fuoco le proiezioni reali (secondo scale temporali differenti) del fenomeno. I mondi della “finzione” e dell’ordine giuridico si incontrano, collaborano, il primo mostra una straordinaria capacità di anticipare la realtà e gli scenari futuri<sup>10</sup> (aiuta a capire problemi che non possono essere studiati guardando il passato)<sup>11</sup>, la “provocazione” diventa sempre più verosimile, fino al punto da entrare nell’edificio degli atti normativi<sup>12</sup>.

Parafrasando il Borges di *Tlon, Uqbar, Orbis Tertius*<sup>13</sup>, sembra quasi di vedere l’immagine –in quel racconto splendidamente descritta- di una realtà che comincia a cedere in più punti, di fronte ad un mondo illusorio e fantastico che improvvisamente comincia ad irrompere nel mondo (apparentemente?) reale. Recentemente, anche Alessandro Baricco, nel suo *The Game*<sup>14</sup>, gioca con la contrapposizione/integrazione tra mondo “reale” e oltremondo digitale. Per Baricco, l’habitat dell’uomo digi-

<sup>6</sup> Criticata da Bernard de Fontenelle (avvocato ed aforista francese a cavallo tra ‘600 e ‘700) con la celebre riflessione sui cani e gli orologi (riportata in S. HUSTVEDT, *Le illusioni della certezza*, Torino, 2018, 191): «Dite che le bestie sono macchine così come lo sono gli orologi? Mettete un cane macchina e una cagna macchina fianco a fianco, e alla fine il risultato sarà una terza piccola macchina, mentre due orologi giaceranno fianco a fianco per tutta la vita senza mai produrre un terzo orologio».

<sup>7</sup> Ne parla S. HUSTVEDT, nel suo splendido libro *Le illusioni della certezza*, Torino, 2018, 164 ss.

<sup>8</sup> Nel suo *L’ordine giuridico delle macchine*, Torino, 2003.

<sup>9</sup> Ne parla GERARD M. EDELMAN, *Sulla materia della mente*, Milano, 1993, 293.

<sup>10</sup> Come scrivono V. GALLESE, M. GUERRA, *Lo schermo empatico. Cinema e neuroscienze*, Milano, 2015, 85, «il cinema è visione del mondo, in tutta la sua composita ampiezza e in tutta la sua affascinante imprevedibilità, visione del mondo che allarga i nostri orizzonti percettivi, cognitivi, immaginativi e che si cala talmente a fondo nel dialogo con il suo tempo e con la storia da restituircene un’interpretazione e una modellazione incredibilmente profonda e completa, che va dal nostro rapporto con la tecnica alla nostra vita sociale, culturale e politica».

<sup>11</sup> Cfr. A. MAYNARD, *Films from the future. The technology and morality of sci-fi movies*, Miami, 2018.

<sup>12</sup> Secondo J. BALKIN, *The Path of Robotics Law*, 60, «these literary and artistic sources ... already give us a sense of what robots might do and the problems that their introduction into society might cause». Più in generale, per S. ALMONG, *Literature alongside law as contemporary paradigm*, in *Cultural Dynamics*, 2001, 59, «both law and literature (are) social practices in constant interaction. It contends that aesthetics and poetics are an inextricable part of law, just as legal and ethical structures are an essential part of aesthetic expression. Narrative, as a basic component of human experience, continually provide one of the formative as well as change-inducing forces of cultural activity, including legal activity».

<sup>13</sup> Pubblicato in *Finzioni*, ora in *Tutte le opere*, vol. I, Milano, 1984, 623-641.

<sup>14</sup> A. BARICCO, *The Game*, cit., 88-89, e 191, dove si legge: «il traffico con l’oltremondo, in entrata e in uscita, è diventato immenso e velocissimo, tanto immenso e tanto veloce che spesso conservare una vera linea di demarcazione tra mondo e oltremondo è diventato impossibile, e quasi sempre inutile». Cfr. anche, sulla giustapposizione tra mondo reale e mondo digitale (‘datasfera’, “fatta di miliardi di miliardi di dati”) V.ZENO-ZENCOVICH, *La “datasfera”. Regole giuridiche per il mondo digitale parallelo*, in L. SCAFFARDI (ed.), *I “profili” del diritto. Regole, rischi e opportunità nell’era digitale*, Torino, 2018, 99 ss.

tale «è un sistema di realtà a doppia forza motrice, dove la distinzione tra mondo vero e mondo virtuale decade a confine secondario, dato che l'uno e l'altro si fondono in un unico movimento che genera, nel suo complesso, la realtà. [...] è un sistema in cui mondo e oltremondo girano uno nell'altro producendo esperienza, in una sorta di creazione infinita e permanente, (...) ricamando una trama che legittimamente chiamiamo REALTÀ»<sup>15</sup>.

Emblematico il richiamo che la Risoluzione del Parlamento Europeo del 16 febbraio 2017 contenente *Norme di diritto civile sulla Robotica* fa, in uno dei suoi *Considerando*, alle leggi di Asimov<sup>16</sup> come base di riferimento etica e deontologica dei comportamenti dei ricercatori che operano nel settore della robotica, e del loro impegno a rispettare i principi di beneficenza (secondo cui i robot devono agire nell'interesse degli esseri umani), di non malvagità (in virtù della quale i robot non devono fare del male ad un essere umano), di autonomia (vale a dire la capacità di adottare una decisione informata e non imposta sulle condizioni di interazione con i robot), di giustizia (nel senso di un'equa ripartizione dei benefici associati alla robotica, e di accessibilità economica ai robot addetti all'assistenza a domicilio e alle cure sanitarie<sup>17</sup>).

Invero, le leggi di Asimov possono essere semplicemente una primissima (e parziale) impostazione del discorso sui principi etici e giuridici dell'AI. Il quadro da costruire è assai più complesso, e di questa complessità comincia ad avvertirsi una consapevolezza via via più profonda nei documenti "paranormativi" (*guidelines, statement, studi preparatori*) elaborati ai più vari livelli istituzionali. Si lavora essenzialmente in chiave di adattamento interpretativo dei principi consolidati del costituzionalismo alle applicazioni di AI: l'obiettivo è quello di configurare un'AI «*human-centric*» and «*trustworthy*», appunto perché connotata da un "ethical purpose", e da un obbligo di conformità ai valori fondamentali della convivenza civile (rispetto della dignità umana e dei diritti, eguaglianza e non discriminazione, non interferenza rispetto ai processi democratici, sicurezza, rispetto per la *privacy*), e al benessere delle persone<sup>18</sup>.

<sup>15</sup> Di una "natura digitale delle cose", come ordine spontaneo in cui «un insieme di gruppi sociali ... si connettono per via evolutiva, e competitiva, dandosi regole private che cessano però di essere squisitamente contrattuali, divenendo anche sociali e istituzionali ...», v. A. VENANZONI, *Intersezioni costituzionali – Internet e Intelligenze artificiali tra ordine spontaneo, natura delle cose digitale e garanzia dei diritti fondamentali*, in *Forum di Quaderni Cost.*, 27 aprile 2018, 8.

<sup>16</sup> Nel Considerando T si legge: «considerando che le leggi di Asimov devono essere considerate come rivolte ai progettisti, ai fabbricanti e agli utilizzatori di robot, compresi i robot con capacità di autonomia e di autoapprendimento integrate, dal momento che tali leggi non possono essere convertite in codice macchina». Le famose leggi di Asimov sono le seguenti: «(1) Un robot non può recar danno a un essere umano né può permettere che, a causa del proprio mancato intervento, un essere umano riceva danno. (2) Un robot deve obbedire agli ordini impartiti dagli esseri umani, purché tali ordini non contravvengano alla Prima Legge. (3) Un robot deve proteggere la propria esistenza, purché questa autodifesa non contrasti con la Prima o con la Seconda Legge (cfr. Isaac Asimov, *Circolo vizioso*, 1942), e (0) Un robot non può recare danno all'umanità, né può permettere che, a causa del proprio mancato intervento, l'umanità riceva danno».

<sup>17</sup> Come sottolinea infatti A. BARICCO, *The Game*, Torino, 2018, 320, «... il Game è difficile. Magari divertente, ma troppo difficile. (...) Per sopravviverci bisogna avere skill non indifferenti che peraltro non vengono insegnate. (...) Non ci sono reti di protezione, né sistemi per recuperare chi è caduto. (...) Non lasceremo nessuno indietro non è una frase da Game».

<sup>18</sup> Sul punto, v. DEG, cit., iv. A questi principi di riferimento è legato anche il criterio secondo cui «AI systems should be developed and implemented in a way that protects societies from ideological polarization and algorithmic determinism» (cit., 9), nonché una serie di altri requisiti di una AI 'trustworthy', tra cui: 1. *Accountability*,

La dignità umana viene tradotta, in un significato relazionale, come diritto delle persone umane di sapere *se e quando* stanno interagendo con una macchina o con un altro essere umano, e di decidere *se, come e quando* attribuire determinati compiti ad un sistema artificiale autonomo o ad una persona. L’autonomia delle persone a contatto con AI agents o systems deve comportare altresì che i meccanismi di funzionamento di questi sistemi siano trasparenti e prevedibili, meglio ancora “esplicabili”, «comprehensible and intelligible by human beings at varying levels of comprehension and expertise»<sup>19</sup> (su questo torneremo anche più avanti).

Il bene primario dell’integrità e della sicurezza personale assume una connotazione “emozionale”, mentale, legata all’interazione uomo-macchina, e agli effetti che le diverse modalità che questa interazione può assumere possono riflettersi sull’atteggiamento delle persone, sul modo di considerare i sistemi agenti in particolare quando adottano ‘fisicità’ antropomorfe, sulle influenze distorsive che si possono produrre anche sul piano emotivo, soprattutto per le persone che si trovano in una condizione di debolezza e di vulnerabilità.

Infine, rispetto agli sviluppi dell’AI viene richiamata (sempre nello Statement del Gruppo di esperti della Commissione Europea) un’esigenza di sostenibilità (Simoncini preferisce il concetto più *hard* di precauzione costituzionale), di mantenimento nel tempo (e di garanzia per generazioni future) delle condizioni di base della vita umana e della natura: un’ulteriore conferma, se ce ne fosse bisogno, della preoccupazione che accompagna l’evoluzione di questa tecnologia, pur nella convinzione delle tante opportunità positive che da essa possono derivare.

## 2. AI, Robotics, mondo “digitale”

Ad ogni modo, l’AI sembra davvero qualcosa in più di un set di risorse tecnologiche aggiuntive e avanzate, con le quali fare semplicemente meglio e in modo più rapido ed efficace quello che possiamo già fare. Alcuni aspetti di questo imponente fenomeno ci pongono davanti ad interrogativi etici ed antropologici che rappresentano un vero e proprio cambio di paradigma<sup>20</sup>, al fondo del quale potrebbe esserci una nuova idea (e una nuova esperienza) dell’umanità<sup>21</sup>.

---

secondo diversi meccanismi di compensazione o di ‘reconciliation’ anche non monetari; 2. *Data governance*, in modo da evitare o prevenire il più possibile biases, errori, effetti discriminatori, o da migliorare i sistemi attraverso la ricerca e la verifica degli errori; 3. *Design for all*, nel senso di massima accessibilità ed usabilità di questi *tools*, senza riguardo a condizioni personali, età, disabilità; 4. *Governance of AI autonomy (Human Oversight)*, commisurata in modo crescente proprio al grado di autonomia operativa delle macchine; 5. *Respect for privacy*, in totale compliance con le norme del GDPR; 6. *Robustness*, intesa come sicurezza degli algoritmi durante le diverse fasi di implementazione, e resilienza rispetto ad eventuali attacchi informatici.

<sup>19</sup> Cfr. DEG, cit., 10 ss.; in particolare, secondo questo documento, «*Explicability also requires accountability measures be put in place. Individuals and groups may request evidence of the baseline parameters and instructions given as inputs for AI decision making (...) by the organizations and developers or an AI system, the technology implementers, or another party in the supply chain*».

<sup>20</sup> A. PUNZI, *L’ordine giuridico delle macchine*, Torino, 2003, 18.

<sup>21</sup> Nelle Draft Ethics Guidelines for Trustworthy AI, predisposto dall’High Level Expert Group on Artificial Intelligence della Commissione Europea (d’ora in avanti DEG), del 18 dicembre 2018, si legge che “*Artificial Intelligence is one of the most transformative forces of our time, and is bound to alter the fabric of society*”. In termini, v. anche F. DE VANNA, *Diritto e nuove tecnologie: il nodo (controverso) della regolazione giuridica*, in *Lo Stato*, 2018, n. 11, 388.

Inoltre, come è stato acutamente sottolineato, l'AI non è un problema temporaneo, non è qualcosa che ad un certo punto sparisce, ma è una novità destinata a provocare una fase nuova dell'esperienza umana con la quale dovremo abituarci a convivere nel tempo<sup>22</sup>.

Intanto, non è facile stabilire cosa è AI; manca una definizione univoca e generalmente accettata, e soprattutto cambia nel tempo<sup>23</sup>, il che è normale per una tecnologia 'generativa', che «changes over time as people work with and through new technology»<sup>24</sup>.

Di certo, non è solo robotica<sup>25</sup>. Quando parliamo di sistemi di AI, può esserci un 'corpo' (vale a dire un contenitore fisico, anche variamente antropomorfo) oppure no: questi sistemi possono lavorare in uno spazio fisico ovvero in uno spazio virtuale.

È possibile pensare a pezzi di AI anche dentro di noi (interfacce neurali, *Brain computer interfaces*), come risorse di potenziamento cognitivo; e i dilemmi non sono meno urgenti e difficili di quelli che riguardano l'*enhancement* genetico e neurobiologico<sup>26</sup>.

D'altronde questo vale anche per la parola e la nozione di 'robot' in senso stretto. Come scrive J. Engelberger, «I can't define a robot, but I know one when I see one». In altri termini, anche un robot, che a sua volta è una proiezione parziale del più ampio *field* dell'AI, può essere tante cose e può assumere diverse configurazioni tassonomiche, che dipendono da cosa è chiamato a fare, dall'"ambiente" di lavoro, dalle tipologie di interazione con gli esseri umani, dai gradi o livelli di autonomia decisionale e/o operativa.

Dunque, AI appare effettivamente un "*umbrella term*", una parola "contenitore", che include –si può dire in via di prima approssimazione– una varietà di tecniche computazionali e di processi associati (di tipo algoritmico) dedicati a migliorare l'abilità delle macchine nel fare cose che richiedono intelligenza<sup>27</sup>.

Sullo sfondo, e anzi al cuore di questo processo, scorre un flusso inesauribile e gigantesco di dati. Lo spiega, con la consueta profondità di pensiero, e con qualche riflesso di inquietudine, Roberto Calasso, nel suo recente libro *L'innominabile attuale*. Calasso parla di contenuti informativi che si moltiplicano e si espandono senza tregua e in ogni direzione, e che si rivelano alla fine autosufficienti, capaci di espandersi senza ricorrere ad alcunché di esterno (per capirci l'esterno saremmo noi e la capacità

<sup>22</sup> Cfr. L. ALEXANDRE, *La guerra delle intelligenze*, cit., 295, sottolineando che «l'IA ormai non è più un'opzione disattivabile, un interruttore che avremo ancora la possibilità di spegnere. È diventata indispensabile». Sempre questo A. ha messo in evidenza che il cervello (tra neuroscienze e AI) è diventato il principale campo di battaglia di questo secolo.

<sup>23</sup> R. ANGELINI, *Intelligenza artificiale e governance. Alcune riflessioni di sistema*, in *Astrid*, n. 14/2017, 1. Una definizione ampia si legge in DEG, iv: «Artificial Intelligence refers to systems designed by humans that, given a complex goal, act in the physical and digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to learn to adapt their behaviour by analysing how the environment is affected by their previous actions».

<sup>24</sup> Per questa definizione di 'generative technology', cfr. J. BALKIN, *The Path of Robotics Law*, 47.

<sup>25</sup> M. LUCIANI, *La decisione giudiziaria robotica*, in *Nuovo diritto civile*, n. 1/2018, 7.

<sup>26</sup> Sul tema, sia consentito il rinvio ad A. D'ALOIA, *Oltre la malattia: metamorfosi del diritto alla salute*, in *Biolaw Journal*, n. 1/2014, 87 ss., 93 ss.

<sup>27</sup> Per questa definizione, v. il Report di AA.VV., *Artificial Intelligence & Human Rights: Opportunities and Risks* (d'ora in avanti AIHR), Berkman Klein Center for Internet & Society at Harvard University, 25 sept 2018, 10, dove si sottolinea altresì che «what is considered artificial intelligence changes over time».

umana di controllare e gestire questi flussi informativi e computazionali): «sono loro, i Big data, che pensano e amministrano coloro da cui hanno avuto origine»<sup>28</sup>. Gli fa eco Paolo Zellini, un autorevole scienziato dell’analisi numerica, che scrive, nel suo ultimo libro (*La dittatura del calcolo*<sup>29</sup>), «Semplicemente noi non possiamo disporre di tutta l’informazione di cui dispongono gli algoritmi. Forse è già successo l’irreparabile, e il criterio più diffuso per confrontarci con la macchina non poggia tanto su una presunta idea di coscienza o di libertà, che nessuno sa definire, quanto su un’imperiosa categoria di efficienza, a tutto vantaggio della macchina...».

### 3. Perché il diritto deve occuparsi di AI. Dalla cibernetica giuridica al diritto per i sistemi che imparano da soli

Perché il diritto deve interessarsi a queste cose? È chiaramente una domanda retorica, la risposta è scontata, il diritto è semplicemente travolto, sfidato da queste cose, e non è solo una difesa d’ufficio di questo focus.

La scienza giuridica ha sempre seguito l’andamento ciclico dell’interesse per l’AI; penso, per rimanere al dibattito italiano, ai lavori pionieristici di Mario Losano<sup>30</sup>, Vittorio Frosini<sup>31</sup>, Stefano Rodotà<sup>32</sup>, degli anni ‘60 e ‘70, anche se era soprattutto cibernetica applicata al diritto, giurimetria, tecnologia strumentale ad una più efficiente amministrazione della giustizia<sup>33</sup>, piuttosto che comprensione o consapevolezza dell’impatto modificativo che un’AI avrebbe potuto avere sugli oggetti fondamentali della riflessione giuridica, e in particolare sui diritti dell’uomo.

Per Losano, «la giurimetria si occupa di materie come l’analisi quantitativa del comportamento giudiziale, dell’applicazione della teoria della comunicazione e dell’informazione all’espressione giuridica, dell’uso della logica nel diritto, del reperimento di dati giuridici con mezzi meccanici ed elettronici e della formulazione d’un calcolo della probabilità applicato all’attività giudiziaria. [...] la *jurisprudence* è un’attività razionale; la giurimetria è uno sforzo per utilizzare i metodi della scienza nell’ambito del diritto. Le conclusioni cui perviene la *jurisprudence* sono semplicemente discutibili; le conclusioni della giurimetria sono invece verificabili. La *jurisprudence* medita su essenze, fini e valori; la giurimetria studia metodi d’indagine».

L’apparente riduzione del campo lascia però già intravedere alcune intuizioni prospettiche. Losano capisce che la cibernetica andrà oltre la giurimetria man mano che si affinerà la disponibilità di risorse tecnologiche<sup>34</sup>. Del resto, già alcuni dei settori di indagine giurimetrica apparivano idonei a riflettere interrogativi più profondi e ulteriori rispetto ad una semplice esigenza o modalità di raccolta, conservazione, elaborazione e rielaborazione di dati elettronici rilevanti per il diritto.

<sup>28</sup> R. CALASSO, *L’innominabile attuale*, Milano, 2017.

<sup>29</sup> P. ZELLINI, *La dittatura del calcolo*, Milano, 2018, 121, 157.

<sup>30</sup> M. G. LOSANO, *Giuscibernetica: macchine e modelli cibernetici nel diritto*, Torino, 1969.

<sup>31</sup> V. FROSINI, *Cibernetica, diritto e società*, Milano, 1973.

<sup>32</sup> S. RODOTÀ, *Elaboratori elettronici e controllo sociale*, Bologna, 1973.

<sup>33</sup> Si veda anche G. DI FEDERICO, *L’uso di strumenti elettronici nell’amministrazione della giustizia*, in Riv. Trim. dir. proc. Civ., 1966/2, 626.

<sup>34</sup> M. LOSANO, *op.cit.*, 93 e 185.

A parte la citazione iniziale di Aurel David, che, parlando del suo libro su *Le cybernétique et l'humain*<sup>35</sup>, afferma «questo libro sarà consacrato al rumore della caduta di buona parte dell'uomo nella materia ed al rumore che la materia fa penetrando impetuosamente nella nostra metafisica e nella nostra morale», Losano riprende, da un lato, le riflessioni di Lee Lovinger<sup>36</sup>, secondo cui la definizione dei confini della giurimetria «di certo di modificherà ed estenderà, man mano che esperimenti ed esperienze risolveranno problemi specifici»; dall'altro, le preoccupazioni di Joseph J. Spengler<sup>37</sup> sul pericolo che l'elaboratore elettronico possa diventare uno strumento di ulteriore asservimento dell'individuo allo Stato, e di Walter Berns<sup>38</sup> sulla “equivoca neutralità valorativa” di tali possibilità tecnologiche<sup>39</sup>.

Anche lo studio di Alfonso Masucci sull'atto amministrativo informatico<sup>40</sup> (siamo agli anni '90) si muoveva in questa logica adattiva e gradualistica, cioè capire come le nuove possibilità tecnologiche, e in particolare la digitalizzazione della P.A., avrebbero potuto riversarsi sugli schemi classici dell'atto amministrativo, a cominciare dal tema della discrezionalità e dei suoi differenti “gradi”.

Oggi lo scenario è completamente cambiato. L'evoluzione del *machine learning* ci pone di fronte ad un contesto in cui la tradizionale (per quanto progressivamente sofisticata) programmazione della macchina viene sostituita dalla sua ‘educazione’ ad imparare da sola<sup>41</sup>, e a trarre da ciò che impara, in modo sempre più autonomo<sup>42</sup>, conseguenze in termini di scelte, azioni, operazioni<sup>43</sup>.

Baricco ricorda il caso di Alpha Go che ha battuto il campione mondiale (umano) di Go nel 2016<sup>44</sup>: un *software* allenato facendogli memorizzare decine di milioni di partite e di mosse di quel gioco, ma appunto, «il bello iniziò quando i programmatori si misero a lavorare con reti neurali profonde (...) ottenendo un risultato che ha, questo sì, del favoloso: Alpha Go impara dagli umani ma poi fa di testa sua, inventa mosse che gli umani non hanno mai fatto, applica strategie che nessun umano aveva pensato...». Potremmo dire, con un incrocio di parole, che quando si arriva a questo punto, il gioco

<sup>35</sup> Paris, 1965, 30.

<sup>36</sup> L. LOVINGER, *Jurimetrics: the methodology of legal inquiry*, in *Jurimetrics*, London-New York, 1963, 8.

<sup>37</sup> J. J. SPENGLER, *Machine-Made Justice: Some Implications*, in *Jurimetrics*, London-New York, 1963, 44 ss.

<sup>38</sup> W. BERNS, *Law and behavioral science*, *ibid.*, 201 ss.

<sup>39</sup> I riferimenti bibliografici e le citazioni sono riprese da M. LOSANO, *Giuscibernetica*, cit. 93, e 101.

<sup>40</sup> A. MASUCCI, *L'atto amministrativo informatico*, Napoli, 1993.

<sup>41</sup> L. ALEXANDRE, *La guerra delle intelligenze*, cit., 17.

<sup>42</sup> Così spiazzando l'affermazione dei primi studiosi delle macchine (come Charles Babbage, nella prima metà dell'ottocento), riportata in J. BERNSTEIN, *Uomini e macchine intelligenti*, cit., 179, secondo cui «*La macchina analitica non ha la pretesa di creare alcunché di originale. Essa può fare tutto ciò che noi siamo in grado di ordinarle di fare*». Sempre J. BERNSTEIN (202) ricorda la frase di una conversazione tra McCarthy e Minsky, indicativa del vertiginoso livello di incertezza e di imprevedibilità che caratterizza gli scenari possibili di sviluppo della tecnologia dell'AI: «*Quando arriverà a quel punto, credo non sapremo più come lo fa*».

<sup>43</sup> Per una definizione di autonomia, nel senso che «*smarter systems ... develop and can perform tasks independently from human operators and without human control*», v. Statement, cit., 7.

<sup>44</sup> A. BARICCO, *The Game*, cit., 203. Anche M. TEGMARK, *Vita 3.0. Essere umani nell'era dell'intelligenza artificiale*, ed. it., Milano, 2018, 121, evidenzia che Alpha Go ha fatto una mossa molto creativa, «in contrasto con millenni di saggezza umana, che circa 50 mosse più tardi si è rivelata cruciale per sconfiggere Lee Sedol, leggenda del Go», aggiungendo significativamente questa riflessione: «*Perché è stata una cosa tanto importante per me personalmente? Be', ho confessato prima che considero intuizione e creatività due dei miei tratti umani fondamentali e, come adesso spiegherò, ho avuto la percezione che Alpha Go presentasse entrambi*».



non è semplicemente giocato, ma, a sua volta, “gioca”: per usare l’espressione di Hofstadter<sup>45</sup>, la mente del programma si distingue da quella del suo programmatore<sup>46</sup>.

Ma chi educa le macchine? Con quali norme etiche? Quali sono o possono essere i principi comuni in un tempo caratterizzato da un esasperato pluralismo etico? Come si può evitare o controllare il rischio di hackeraggio dei sistemi agenti autonomi?<sup>47</sup> Quale efficacia regolativa concreta conserva ancora il “codice” (cioè la legge autonoma della rete, la sua architettura normativa spontanea nella visione di Lessig) nel contesto dei “*self-learning systems*” e del “comportamento emergente” delle macchine e dei sistemi artificiali agenti?<sup>48</sup>

Responsabilità, giustizia, salute, lavoro, *privacy*, decisione autonoma, sono solo alcuni dei punti di ricaduta sul diritto come esperienza (e tecnica di regolazione) sociale di questa complessa categoria dell’evoluzione tecnologica. Come si può vedere, si tratta di *assets* fondamentali del discorso giuridico; in questo senso, si è aperto ormai un campo di lavoro straordinariamente affascinante, in larga parte ancora da costruire e da esplorare, sullo sfondo del quale si colgono sfide etiche e sociali davvero molto intense e al tempo stesso incerte nell’esito, a maggior ragione se proviamo ad incrociare lo sviluppo di AI systems con altri temi anch’essi in forte movimento e certamente interconnessi, come biotecnologie, nanotecnologie, neuroscienze.

L’interesse del diritto per l’AI trova allora una giustificazione quasi “naturale”, intrinseca. Il diritto è, in una prospettiva realista, un prodotto di forze sociali<sup>49</sup>, il riflesso di esigenze e problemi della vita sociale contemporanea; l’AI è ormai, e sempre più lo sarà, un elemento essenziale nella vita delle persone, nel sistema di relazioni sociali, che è proprio ciò che il diritto è chiamato a mediare con le sue regole e le sue risorse.

#### 4. Segue: Macchine (sempre più) autonome e responsabilità

In secondo luogo, robots e sistemi di AI possono creare danni fisici o morali o economici con il loro comportamento<sup>50</sup>, che sempre più diventa meno prevedibile o programmato<sup>51</sup>. Non è facile calcolare i possibili effetti negativi, i modi dell’interazione di questi sistemi con l’ambiente circostante e con gli interessi e le esigenze umane.

La questione non è solo quella delle *driverless cars* e delle scelte di autonomia “operativa” in situazioni eticamente complesse (si pensi al complicato “dilemma delle collisioni inevitabili”, di cui parlano Amoroso e Tamburrini nel loro contributo). L’uso militare dei sistemi di AI, e quindi la diffusione di

<sup>45</sup> In *Godel, Escher, Bach: un’eterna ghirlanda brillante*, trad. it., Milano, 1984, 655.

<sup>46</sup> Cfr. anche C. ACCOTO, *Il mondo dato*, cit., 71, secondo cui «la rivoluzione del machine learning arriverà ad automatizzare l’automazione, un’automazione al quadrato».

<sup>47</sup> Su quest’ultimo profilo, v. L. ALEXANDRE, *La guerra delle intelligenze*, cit., 13. Sui rischi del machine learning anche per la trasparenza degli ‘automated systems’, v. *Algorithms and Human Rights*, cit., 37-38.

<sup>48</sup> È la preoccupazione espressa da J. BALKIN, *The Path of Robotics Law*, 52.

<sup>49</sup> Lo ricorda, richiamando l’insegnamento di Oliver Wendell Holmes Jr., J. BALKIN, *The Path of Robotics Law*, in *California Law Review*, vol. 6, 2015, 46.

<sup>50</sup> R. CALO, *Robotics and the Lessons of Cyberlaw*, in *California Law Review*, vol. 103, 2015, 513 ss., 515, 532.

<sup>51</sup> R. CALO, *ulc*, parla di “emergent behavior”, per sottolineare l’incremento dei livelli di autonomia operativa dei sistemi robotici e di AI.

armi autonome<sup>52</sup>, mette sotto stress alcune fondamentali clausole generali del diritto internazionale umanitario, come il principio di distinzione (tra combattenti nemici e obiettivi militari da un lato, e popolazione civile e oggetti civili dall'altro), il principio di proporzionalità e quello di precauzione, a parte il profilo –segnalato sempre dai due autori prima richiamati– dell'incompatibilità con il principio della dignità umana della possibilità stessa che un agente artificiale possa adottare e mettere in pratica decisioni e comportamenti in grado di offendere la vita, l'integrità fisica, i beni di persone umane.

Ovviamente, quando si parla di responsabilità, danni, distribuzione dei diritti e degli obblighi legati a determinati comportamenti e scelte, siamo su un terreno usuale per il diritto come teoria e come scienza pratica che cerca di risolvere problemi, di regolare conflitti. Nondimeno, le categorie consolidate subiscono una pressione distorsiva notevole, non sembrano sempre adattabili; l'incrocio con la novità dei sistemi di AI lascia sul terreno fratture concettuali che faticano a comporsi.

Non è sempre facile definire *cosa* imputare *a chi*, capire *chi* e *come* è responsabile per i danni derivanti dalle decisioni o dalle operazioni di un *AI system*, o di un *AI agent*, in particolare quando l'azione presenta livelli elevati di autonomia<sup>53</sup> ovvero quando implica scelte e opzioni che hanno un contenuto e delle conseguenze moralmente apprezzabili o valutabili (si pensi al caso, prima richiamato della macchina a guida autonoma che viene a trovarsi in una situazione di incidente inevitabile, cd. *"lose-lose situation"*).

In questa ottica, non sembra molto rilevante ragionare sul tipo di autonomia che la macchina può esprimere. Non c'è dubbio che quella dei sistemi di AI sia, almeno allo stato, un'autonomia meramente "operazionale" (così si esprimono anche Tamburrini e Amoroso nel loro contributo), o "puramente tecnologica" (come afferma la Risoluzione del Parlamento europeo)<sup>54</sup>. Il punto però è che c'è una decisione presa al di fuori di una sequenza programmatica controllata o almeno influenzata: e questa decisione può avere implicazioni che –fuori dal contesto AI– definiremmo morali.

Rispetto alla decisione algoritmica o dell'agente artificiale autonomo possono giocare una serie di fattori causativi o derivativi, e la catena di controllo sul modo di operare degli *algorithmic tools* non è sempre chiara e precisamente descrivibile nella distribuzione (tra programmatori, sviluppatori, *users*) di compiti e capacità di inferenza.

Insomma, possiamo concordare sulla necessità di un "meaningful Human Control", nel senso che «humans – and not computers and their algorithms – should ultimately remains in control»<sup>55</sup>; ma non possiamo escludere che in alcune situazioni questo principio possa mostrare delle crepe.

<sup>52</sup> Sulla robotica militare e sui droni come 'killing machines', v. altresì C. SALAZAR, *Umano, troppo umano ... o no? Robot, androidi e cyborg nel "mondo del diritto" (prime notazioni)*, in *Biolaw Journal*, n. 1/2014, 273 ss.

<sup>53</sup> A. VENANZONI, *Intersezioni costituzionali – Internet e Intelligenze artificiali*, cit., 4, ricorda che nel marzo 2018, in Arizona, una *self-driving car* di Uber ha travolto e ucciso un pedone. Si tratterebbe, a quanto viene riferito, del primo incidente mortale causato da un'auto a guida autonoma.

<sup>54</sup> Anche per Statement, cit., 9, «*Autonomy in the etically relevant sense of the word can therefore only be attributed to human beings. It is therefore somewhat of a misnomer to apply the term 'autonomy' to mere artefacts, albeit very advanced complex adaptive or even 'intelligent' systems. (...) autonomy in its original sense is an important aspect of human dignity that ought not to be relativised*».

<sup>55</sup> Statement, cit., 10. Secondo questo documento, inoltre, «*moral responsibility, in whatever sense, cannot be allocated or shifted to 'autonomous' technology*».

Per gli *Automated vehicles*, com’è noto, la legge tedesca approvata l’anno scorso conferma lo schema della responsabilità del guidatore, che deve comunque essere presente sull’auto, pur potendo affidare in determinate situazioni, e temporaneamente, la conduzione del veicolo al sistema automatico. La normativa prevede l’obbligo di una ‘scatola nera’ che registri le informazioni relative alla guida in modo da rendere possibile una diversa ricostruzione delle responsabilità in caso di incidente, e un coinvolgimento del costruttore-produttore e/o del programmatore del sistema.

È una soluzione provvisoria. Come spiega molto bene R. Calo<sup>56</sup>, «the story of robots and the law is only beginning, (...) and we will only understand the true impact of robots on American Law and legal institutions in the fulness of time».

Il problema che riusciamo adesso solo ad intuire è quello del “comportamento emergente”, vale a dire l’abilità di un sistema “autonomo” di comportarsi in «complex, unanticipated ways»<sup>57</sup>. Quando il livello di autonomia di certe operazioni, e nello specifico della guida di veicoli su strada, diventerà più consistente e non strettamente riconducibile alla programmazione, bisognerà porsi il problema di come decide la macchina, di quali sono i criteri che orientano le sue scelte, e se è eticamente e giuridicamente ragionevole che ad essere responsabile sia sempre e solo l’operatore finale, vale a dire il proprietario o il guidatore del veicolo.

Probabilmente continuerà ad essere utilizzato lo schema della *liability* del proprietario/guidatore, temperato dalla possibilità di ricerca di difetti specifici di produzione o di programmazione. In fondo, il diritto può prendere le strade più semplici sul piano pratico, e questa indubbiamente lo è. Anche rispetto ad un veicolo con un grado di autonomia e di “*emergence*” crescente, l’acquisto e l’uso con la consapevolezza delle caratteristiche prima richiamate può fungere da giustificazione razionale dell’imputazione di responsabilità<sup>58</sup>.

Analogamente, per gli AI *agents* o *systems* utilizzate da amministrazioni pubbliche, è normale che a rispondere di eventuali danni non previsti sia comunque il soggetto pubblico o l’Autorità ‘titolare’ della funzione ‘delegata’ ai (o svolta con la collaborazione dei) sistemi autonomi.

Il nodo vero, che mostra tutta la sua complessità nelle situazioni che abbiamo chiamato “lose-lose”, riguarda il profilo stesso del tipo di decisione che la macchina può assumere, e dei presupposti morali che possono supportarla. Possiamo assegnare al sistema “autonomo” dei criteri legali e morali di comportamento? O configurare degli algoritmi che incorporino questa dimensione etica o giuridica? Se sì, quali criteri possiamo o dobbiamo insegnare alla macchina?

<sup>56</sup> R. CALO, *Robots in American Law*, in E. HILGENDORF, U. SEIDEL (eds.), *Robotics, Autonomics, and Law*, Baden-Baden, 2017, 107-108.

<sup>57</sup> R. CALO, o.c., 103.

<sup>58</sup> La tesi invece dell’imputazione della responsabilità nel produttore del veicolo a guida autonoma, è avanzata e sostenuta da C. HOLDER, V. KHURANA, F. HARRISON, L. JACOBS, *Robotics and Law: key legal and regulatory implications of the robotic age*, I, in *Computer Law and Security Review*, 2016, 386. Nella Risoluzione del parlamento europeo del febbraio 2017, invece, si preferisce indicare un modello di distinzione/frammentazione della responsabilità, nel senso che «una volta individuati i soggetti responsabili in ultima istanza, la loro responsabilità dovrebbe essere proporzionale all’effettivo livello di istruzioni impartite al robot e al grado di autonomia di quest’ultimo, di modo che quanto maggiore è la capacità di apprendimento, o l’autonomia di un robot e quanto maggiore è la durata della formazione di un robot, tanto maggiore dovrebbe essere la responsabilità del suo formatore» (p. 56).

Invero, nelle situazioni in cui gli interessi in collisione (inevitabile) sono di livello gerarchicamente disomogeneo sul piano della valutazione etica e giuridica (la vita di una o più persone da un lato, beni materiali dall'altro), non sembrano esserci particolari problemi: il sacrificio del bene di minor valore è sicuramente la scelta più conforme ai valori fondamentali, sia sul piano etico che su quello giuridico<sup>59</sup>.

Quando invece gli interessi hanno lo stesso peso assiologico, è complicato scegliere uno standard decisionale che possa essere eticamente condiviso.

L'approccio del minor numero non sembra invero applicabile quando in gioco c'è la vita delle persone. Ogni vita ha un valore incomparabile: non è scontato fare valutazioni quantitative, dire che la vita vale meno della vita di due persone, o che è possibile sacrificare una vita per salvarne altre in numero maggiore.

Possiamo dire che un soggetto che facesse questo non sarebbe punibile, ma non certamente che sia obbligato a farlo per non essere punito. Ad ogni modo è un criterio molto instabile e tutt'altro che "common". Potrebbe andare bene per il sistema americano, non per quello tedesco dove, come si ricorderà, il Tribunale Costituzionale nel 2006 ha dichiarato illegittima costituzionalmente, per violazione del principio della dignità umana, la norma dell'*Aviation Security Act* che prevedeva l'autorizzazione per le forze aeree tedesche ad abbattere aerei eventualmente caduti nelle mani di terroristi per impedire che siano usati come strumenti di attacco e di morte come è successo l'11 settembre negli USA<sup>60</sup>.

Un altro criterio potrebbe essere quello della decisione affidata ad un "Random generator". In sostanza, la macchina sceglierebbe sulla base di un algoritmo assolutamente casuale, una sorta di lotteria, solo che la posta in gioco è la vita delle persone.

Apparentemente, è un meccanismo perfettamente neutrale, egualitario. Anch'esso presenta tuttavia, come è stato ben evidenziato, alcuni lati oscuri: ad esempio, quando la macchina si sposta dalla sua originale traiettoria potrebbe mettere a rischio un pedone che prima sarebbe stato al sicuro, e che non ha avuto alcun concorso nella configurazione della situazione di incidente inevitabile. In altre parole, il rischio viene riallocato e può coinvolgere altri soggetti rispetto a quelli inizialmente coinvolti<sup>61</sup>.

## 5. AI, diritto, diritti. Opportunità e problemi

L'impatto di questi sistemi e del loro funzionamento su molti diritti fondamentali in diversi settori – che vanno dalla giustizia penale, all'accesso al credito, ai servizi collegati alla salute, alle procedure di reclutamento e assunzione, all'accesso all'università, alla partecipazione politica – appare già ora molto consistente (e ancora di più mostra di poter essere in un futuro più o meno prossimo), sia in senso positivo, vale a dire in termini di opportunità e di chances<sup>62</sup>, sia in termini di possibili rischi e preoc-

<sup>59</sup> J. FELDLE, *Delicate decisions: legally compliant emergency algorithms for autonomous cars*, in E. HILGENDORF, U. SEIDEL (eds.), *Robotics, Autonomics, and Law*, cit., 199.

<sup>60</sup> Cfr. ancora J. FELDLE, *Delicate decisions*, cit., 200-201, nonché E. HILGENDORF, *Automated driving and the Law*, ivi, 190.

<sup>61</sup> Per queste considerazioni, v. J. FELDLE, *Delicate decisions*, 203.

<sup>62</sup> Ad esempio nelle procedure di scoring per l'accesso al credito bancario e finanziario; cfr. AIHR cit., 30 ss.

cupazioni<sup>63</sup>. Non avrebbe molto senso polarizzare il dibattito tra sostenitori o oppositori degli sviluppi in tema di AI<sup>64</sup>. Bisogna invece cercare di avvalersi dei benefici, minimizzando i rischi e le criticità che certamente ci sono<sup>65</sup>.

I sistemi di AI saranno uno dei grandi blocchi dello sviluppo economico mondiale in questo secolo<sup>66</sup>. La sfida è (e sarà) quella di orientare (o almeno di ridurre i contrasti di) questa nuova imponente evoluzione tecnologica ed economica rispetto ai principi di tutela della dignità e della sicurezza umana, e dei diritti fondamentali. Una sorta di attualizzazione del messaggio contenuto nel secondo comma del nostro art. 41 Cost., e più in generale in tutto il disegno costituzionale di società, fondato sulla centralità del lavoro.

Lo sviluppo della robotica e dei sistemi di AI avrà (sta già avendo) una serie di ricadute significative sul mercato del lavoro, sulla stessa sicurezza nei luoghi di lavoro in considerazione delle più diffuse occasioni di coesistenza e di interazione tra umani e agenti “artificiali”<sup>67</sup>.

Non è solo questione di posti di lavoro che si perdono, e che in parte potranno essere compensati da nuove esigenze e da nuove opportunità occupazionali indotte proprio dall’evoluzione tecnologica nel senso dell’automazione<sup>68</sup>.

La sfida dell’AI richiede un aggiornamento complessivo dei modelli di istruzione e formazione professionale, e un ripensamento di alcune strutture portanti del welfare. La «liberazione delle persone dalla monotonia del lavoro manuale» (per usare l’espressione della Risoluzione del Parlamento Europeo

<sup>63</sup> Cfr. AIHR, cit., 17 ss., dove si parla di un “AI’s multifaceted human rights impacts”, sottolineando altresì (a pag. 58) che «*the relationship between artificial intelligence and human rights is complex. A single Ai application can impact a panoply of civil, political, economic, social, and cultural rights, with simultaneous positive and negative impacts on the same right for different people*». Tornando all’esempio delle procedure di accesso al credito, non si può escludere, come è stato notato (AIHR, cit., 31) che «*AI-based decision-making algorithms in the financial sector will adversely impact the freedoms of opinion, expression, and association. In an era where all data is credit data, individuals may feel chilled from expressing certain points of view or associating with others, out of fear that an algorithm may use their behavior against them in the financial context*».

<sup>64</sup> Con grande acutezza, R. CALO, *Artificial Intelligence Policy: A Primer and Roadmap*, in *Univ. of California Law Rev.*, vol. 51, 2017, 431, spiega che «*devoting disproportionate attention and resources to the AI apocalypse has the potential to distract policymakers from addressing AI’s more immediate harms and challenges and could discourage investment in research on AI’s present social impacts*». Anche M. TEGMARK, *Vita 3.0*, 145, a proposito di giustizia ‘robotica’ o ‘algoritmica’, dice: «*Non si tratta di prendere una decisione del tipo ‘tutto o nulla’ in merito ai robogiudici, ma di una decisione sul grado e sulla velocità con cui vogliamo applicare l’IA nel nostro sistema legale*».

<sup>65</sup> Come scrive C. ACCOTO, *Il mondo dato*, cit., 57, «*il dubbio se si tratti di tecnologie assistive (a supporto di umani in difficoltà come bambini, anziani, malati), tecnologie aumentative (per accrescere le capacità umane oltre la loro massima espressione) o tecnologie sostitutive (con lo scopo di eliminare gli umani dall’espletamento di attività) è vivo più che mai*».

<sup>66</sup> Per R. ANGELINI, *Intelligenza artificiale e governance. Alcune riflessioni di sistema*, cit., 10, con l’IA siamo probabilmente di fronte a una nuova rivoluzione industriale, non assimilabile alle precedenti, che cambierà il modello stesso di sviluppo.

<sup>67</sup> Va detto che, come ha evidenziato ancora la Risoluzione del Parlamento Europeo del 2017, la robotica offre una serie di potenzialità per migliorare la sicurezza sul posto di lavoro mediante il trasferimento di alcuni compiti pericolosi e dannosi dagli esseri umani ai robot (par. 46).

<sup>68</sup> Come scrive ancora R. ANGELINI, *Intelligenza artificiale e governance. Alcune riflessioni di sistema*, cit., 15, «*è comunque ragionevole ritenere che determinati lavori e professioni vedranno una decrescita a causa dell’impiego dell’IA, altri invece emergeranno proprio in ragione di esso. Nessuno, verosimilmente, ne sarà completamente immune*».

sulla robotica) è una bella prospettiva, ma a patto di costruire un'alternativa fatta di tante cose concrete, come formazione delle nuove competenze richieste dallo sviluppo della robotica e dell'AI, nella scuola e durante la vita lavorativa, percorsi di reinserimento nel lavoro, reddito di cittadinanza o comunque strumenti compensativi per le situazioni di disoccupazione di lungo periodo. Altrimenti, diventa una strada a senso unico verso la perdita di senso della coesione sociale e, in definitiva, della democrazia<sup>69</sup>.

I contributi qui raccolti evidenziano molto bene questa ambivalenza, di cui mostra piena consapevolezza lo stesso Norbert Wiener, uno dei "padri" della cibernetica. Scrive infatti il matematico statunitense: «Quelli fra noi che hanno contribuito alla nuova scienza della cibernetica si trovano così in una posizione morale a dir poco scomoda. Abbiamo contribuito alla nascita di una nuova scienza che, come ho detto, comporta sviluppi tecnici con grandi possibilità per il bene e per il male. Non possiamo fare altro che consegnarla al mondo che ci circonda, e questo è il mondo di Belsen e di Hiroshima. Non abbiamo neanche la scelta di arrestare questi nuovi sviluppi tecnici. Essi appartengono alla nostra epoca, e il massimo che riusciremmo ad ottenere cercando di sopprimerli sarebbe di metterli nelle mani dei più irresponsabili e venali dei nostri ingegneri. Il meglio che possiamo fare è agire in modo che un vasto pubblico comprenda le tendenze e gli aspetti di questo lavoro»<sup>70</sup>.

## 6. Algoritmi e (rischi di) discriminazioni

L'impatto può essere asimmetrico sul piano sociale. Inoltre, in alcuni settori gli "AI systems" (o *agents*) possono perpetuare (anche indirettamente e involontariamente<sup>71</sup>) modelli e schemi discriminatori, incidendo in modo più severo e penalizzante nei confronti di gruppi minoritari e marginalizzati.

Una serie di studi evidenzia come *software* di previsione dei reati o di calcolo del recidivismo possono basarsi su informazioni che finiscono con l'alimentare una sorta di circolo vizioso, per cui alla fine, ad essere criminalizzata, è la povertà<sup>72</sup>.

E tutto questo in un contesto di opacità e di imperscrutabilità in cui non è facile capire quali sono i meccanismi di funzionamento di questi strumenti, le informazioni sulla base delle quali vengono poi prese o suggerite determinate decisioni.

<sup>69</sup> Cfr. le riflessioni di A. SCHIANCHI, M. FRANCHI, *Democrazia senza*, Parma, 2016.

<sup>70</sup> N. WIENER, *Cybernetics. Or Control and Communication in the Animal and the Machine*, MIT Press, Cambridge, 1948, 54.

<sup>71</sup> Come sottolinea D. CARDON, *Che cosa sognano gli algoritmi*, Milano, 2016, 71, secondo cui «l'algoritmo non ha bisogno di avere un'intenzione discriminatoria per produrre quel genere di effetti discriminatori. Non contiene norme che gli chiedono di individuare le persone nere e le persone bianche. Si accontenta di lasciar fare alle regolarità statistiche che dicono che cognomi e nomi di persone nere sono statisticamente più spesso legati a clic verso ricerche sulle fedine penali. Lasciato a sé stesso, il calcolatore si basa sui comportamenti degli altri internauti e contribuisce, 'innocentemente', se così si può dire, alla riproduzione della struttura sociale, della disparità e delle discriminazioni». Analogamente, v. C. ACCOTO, *Il mondo dato*, cit., 67.

<sup>72</sup> C. O'NEIL, *Armi di distruzione matematica*, Firenze-Milano, 2017, 135 e 143. Secondo questa A., «in un sistema del genere, le persone povere e di colore vengono punite principalmente per il fatto di essere quello che sono e vivere dove vivono».

*Armi di distruzione matematica*, le chiama Cathy O’Neil, e al di là dell’enfasi forse eccessiva della definizione, il problema reale è capire «se abbiamo eliminato il pregiudizio umano o l’abbiamo semplicemente camuffato con la tecnologia». Sta di fatto che alcune domande e informazioni che vengono sottoposte ai soggetti detenuti o anche semplicemente “fermati” in operazioni di controllo e prevenzione del crimine, possono condurre a risultati confermativi o reiterativi di quelle condizioni di debolezza e di difficoltà, alimentando così, più o meno inconsciamente, “un ciclo tossico” di pregiudizi e di parzialità, in cui, «anche se un modello nasce per non tener conto della razza, il risultato è esattamente quello. Nelle nostre città in cui, per la maggior parte, vige una sorta di segregazione razziale, la geografia è un dato che sostituisce perfettamente la razza».

Il problema è innanzitutto di programmazione e “design” dei modelli algoritmici; di come si configura quello che Lawrence Lessig chiama il “codice” del *cyberspace*<sup>73</sup>, la sua architettura che al tempo stesso esprime una normatività autonoma<sup>74</sup>; ma potrebbe non bastare nel mondo del “machine learning”, di fronte a sistemi che apprendono da soli e che prendono ‘autonomamente’ le informazioni di cui hanno bisogno nello sconfinato repertorio informativo del web<sup>75</sup>.

Evidentemente, si pone un tema molto importante di formazione etica (e di promozione di una cultura della cooperazione e della trasparenza) dei programmatori e sviluppatori di *Ai systems* o *agents*, in modo da anticipare conflitti e distanze<sup>76</sup>; e, al tempo stesso, di formazione tecnologica nelle facoltà giuridiche e negli uffici di supporto degli organi istituzionali e amministrativi<sup>77</sup>.

Il diritto ha già conosciuto, e in qualche modo contrastato, le situazioni di discriminazione “indiretta” messe in campo da istituzioni, poteri economici, soggetti privati. Appare molto più complicato educare un algoritmo a perseguire logiche sostanziali di eguaglianza, a capire che in alcuni casi le condizioni di partenza sono storicamente squilibrate, e che hanno bisogno di misure di favore per impedire che le classi svantaggiate finiscano per l’esserlo ancora di più<sup>78</sup>.

<sup>73</sup> Il riferimento è al saggio di L. LESSIG, *The Law of the Horse. What Cyberlaw might teach*, in *Harvard Law Rev.*, 1999, vol. 113, 501 ss.

<sup>74</sup> Con i connessi problemi di responsabilità dei *coders* e le criticità sul piano della trasparenza e dell’accountability; cfr. F. DE VANNA, *Diritto e nuove tecnologie*, cit., 392-393, 398.

<sup>75</sup> I rischi di un algoritmo “strutturalmente incostituzionale” perché discriminatorio sono sottolineati nel contributo di A. SIMONCINI, secondo cui «l’algoritmo predittivo è costruito su un set di dati che è già in partenza discriminatorio. È il principio noto tra i data scientists come GIGO –garbage in garbage out- per cui un algoritmo non può che riflettere la qualità dei dati su cui è costruito».

<sup>76</sup> Insiste molto su questo punto, nella sua visione della precauzione costituzionale, A. SIMONCINI, nel suo contributo.

<sup>77</sup> Così M. TEGMARK, *Vita 3.0*, 146. Anche per R. ANGELINI, *Intelligenza artificiale*, cit., 27, «gli attori pubblici che si confrontano con l’IA dovranno essere preparati secondo una logica multidisciplinare e integrata».

<sup>78</sup> Su questi rischi di un uso discriminatorio degli algoritmi, che possono diventare “new modes of racial profiling”, v. S. UMOJA NOBLE, *Algorithms of Oppression. How Search Engines Reinforce Racism*, New York, 2018, 1 ss., 9, secondo cui «there are several cases that demonstrate how racism and sexism are part of the architecture and language of technology, an issue that needs attention and remediation». Di *algorithmic affirmative actions* parla il Report AIHR, cit., 18.

## 7. AI e processo. Quale spazio per la decisione algoritmica?

L'uso dei *AI systems* nel processo e nelle attività di *crime prevention* (in forte crescita<sup>79</sup>) apre prospettive molto controverse.

Già Losano, nel suo lavoro del 1969, immaginava che la cibernetica potesse essere un fattore di eguaglianza, contro la creatività diseguale e opportunistica del bravo avvocato, l'astuzia che non tutti possono permettersi<sup>80</sup>; e anche solo per il fatto di rendere il processo meno costoso e dunque più accessibile per tutti<sup>81</sup>.

La utilità della decisione giudiziaria "robotica" o "algoritmica" viene declinata soprattutto nel senso della rapidità e della oggettività (potremmo dire anche esattezza della decisione non condizionata da fattori "soggettivi", emozionali, di adeguatezza professionale delle parti e del giudice, finanche di pregiudizi legati al sesso o all'orientamento sessuale, alla razza, alla religione, alla nazionalità, ecc.).

La rapidità è sicuramente un aspetto importante del giusto processo<sup>82</sup>. Il principio della ragionevole durata dei processi esprime da tempo una rilevanza pienamente costituzionale (art. 111 Cost. art. 6 Convenzione EDU).

Ma il diritto e la giustizia sono solo questo? Sul piano dei principi costituzionali non è meno importante l'effettività e la pienezza del diritto alla difesa delle parti, la qualità della decisione giurisdizionale, la capacità del giudice di far emergere la irriducibile peculiarità dei fatti e di calibrare su di essi la decisione, in particolare (o almeno) quando "*at bar*" (per usare l'espressione del celebre libro di S. Jasanoff) arrivano questioni scientifiche o rivendicazioni concrete inedite, non classificabili statisticamente, difficili da collocare in una dimensione standardizzata.

È possibile ricondurre agli schemi astratti della computazione algoritmica la straordinaria varietà dei fatti che il diritto è chiamato a considerare (in modo ragionevole e con proporzionalità<sup>83</sup>), le emozioni, le speranze, il problema della sua irripetibile (e talvolta drammatica) essenzialità che si pone davanti al giudice? In che misura è ipotizzabile, si chiede criticamente M. Luciani<sup>84</sup>, la decisione che risolve con certezza geometrica le questioni di diritto, come ipotizzava Leibniz?

Decidere sulla pericolosità o recuperabilità sociale di un soggetto, se il rischio di recidivismo debba prevalere sulla concessione di una misura di libertà (vigilata o condizionale), può essere affidato (e fino a che punto o misura) ad un meccanismo automatizzato, basato sul trattamento algoritmico di informazioni, precedenti, dati personali o di contesto, che magari possono riflettere una serie di condi-

<sup>79</sup> Come rileva il Rapporto *Algorithms and Human Rights*, preparato da una Commissione di esperti del Consiglio d'Europa (marzo 2018: il titolo completo è "*Study on the Human Rights dimensions of automated data processing techniques (in particular algorithms) and possible regulatory implications*"), 10.

<sup>80</sup> M. G. LOSANO, *Giuscibernetica*, cit., 22.

<sup>81</sup> M. TEGMARK, *Vita 3.0*, cit., 144.

<sup>82</sup> In termini, v. M. LUCIANI, *La decisione*, cit., 4. Aggiunge peraltro M. TEGMARK, *Vita 3.0*, u.l.c., che «*i robogudici potrebbero essere copiati, poiché consisterebbero di poco più che software, consentendo l'esame di tutti i casi pendenti in parallelo anziché in serie*», e questo certamente avrebbe effetti positivi sulla tempestività della risposta giurisdizionale.

<sup>83</sup> Lo ricorda F. DE VANNA, *Diritto e nuove tecnologie: il nodo (controverso) della regolazione giuridica*, cit., 391. Molto acuta appare questa riflessione interrogativa ancora di M. TEGMARK, *Vita 3.0*: «*Se gli accusati vogliono sapere perché sono stati condannati, non dovrebbero avere diritto a una risposta migliore che 'abbiamo addestrato il sistema su una grande quantità di dati, e questo è quel che ha deciso?'*».

<sup>84</sup> M. LUCIANI, *La decisione*, cit., 11, e 23.



zioni in se potenzialmente sfavorevoli e tali da riproporre esiti sostanzialmente confermativi, e perciò discriminatori<sup>85</sup>?

Sono domande che lasciano molti dubbi, anche sulla accettabilità sociale di alcune conseguenze. Massimo Luciani parla, a proposito della decisione giudiziaria robotica, di “effetti di sistema profondi”, ad esempio in termini di vera e propria trasfigurazione di elementi come la difesa tecnica, le argomentazioni dell’avvocato, i ricorsi, la motivazione della decisione giudiziaria<sup>86</sup>.

Molti dei contributi qui pubblicati si confrontano con il complicato rapporto tra AI e struttura del processo (in particolare i saggi di Quattrococo e Simoncini). Il luogo “critico” di questa riflessione è soprattutto il processo penale: del resto, nessun altro sotto-sistema istituzionale ha un impatto potenzialmente più pesante su diritti umani fondamentali, come la libertà personale, il diritto alla sicurezza, la dignità umana, il diritto ad un giusto processo e a non essere considerati colpevoli fino a che non venga accertata la propria responsabilità (nella nostra Costituzione ‘definitivamente’), il diritto del condannato a sperare nella rieducazione e nel reinserimento sociale.

Il *leading case* è il caso Loomis, nel quale la Supreme Court del Wisconsin (confermando la decisione della Corte di grado inferiore) ha legittimato l’uso del software Compas per valutare il rischio di recidiva del condannato, e ha ritenuto rispettato il principio costituzionale del due process, sulla base del fatto che «although the circuit court mentioned the Compas risk assessment, it was not determinative in deciding whether Loomis should be incarcerated, the severity of sentence, or whether he could be supervised safely and effectively in the community».

In altre parole, come rileva Simoncini, l’algoritmo è stato solo uno degli strumenti a disposizione del Giudice per esercitare la sua discrezionalità, ed è servito a supportare e completare altri elementi di valutazione.

Tutto risolto? L’affermazione del principio di ‘non esclusività’ della decisione algoritmica è idoneo a prevenire ogni possibile abuso?

Di sicuro, deve essere un punto fermo, e bene fa il Regolamento europeo a rimarcarlo. È necessario mantenere la natura strumentale, di ausilio, delle tecniche di AI rispetto all’attività del Giudice. Il Giudice è soggetto soltanto alla legge, dice la Costituzione: e questo significa, tra le altre cose, che la legge può essere ‘affiancata’, orientata nell’applicazione da strumenti ulteriori, ma non sostituita mettendo il Giudice di fronte ad automatismi applicativi dipendenti dall’esito di procedure algoritmiche, per quanto alimentate e ‘allenate’ dai dati dei precedenti giurisprudenziali<sup>87</sup>.

Il problema tuttavia, è quello che Simoncini chiama “la travolgente forza pratica dell’algoritmo”: «un indubbio plusvalore pratico connesso alle scelte suggerite automaticamente (dal sistema, dall’algoritmo, dalla profilazione automatica), rispetto alla quale ci si può discostare, ma a patto di impegnarsi in un notevole sforzo (e rischio) valutativo».

---

<sup>85</sup> Per questa preoccupazione, v. ora AIHR cit., 21, che richiama una decisione della Corte Suprema del Canada, nel caso Ewert c. Canada del 13/6/2018, secondo cui «*the use of the assessments may overestimate the risk they pose, continue to cause discrimination, and contribute to unfair prison sentences for Indigenous offenders*», sebbene quest’ultima sia relativa agli strumenti psicologici di *risk assessment*.

<sup>86</sup> M. LUCIANI, *o.u.c.*, 24-25.

<sup>87</sup> M. LUCIANI, *La decisione*, cit., 30, afferma: «*Per essere chiari: se devo scegliere qualcuno di cui non fidarmi, personalmente, scelgo il giudice. E scelgo che sia un essere umano. Umano e consapevole dell’importanza, certo, ma anche dei limiti, della sua funzione*».

La motivazione si conferma allora il cuore delle garanzie costituzionali del processo e della stessa imparzialità del Giudice (ma considerazioni analoghe possono essere fatte anche nei procedimenti amministrativi, come mostra ancora Simoncini nel caso della mobilità degli insegnanti ai sensi della l. 107/2015). Ancora di più nella “società algoritmica” che si va profilando; un modo necessario per bilanciare la tendenziale inaccessibilità dei criteri di funzionamento di questi sistemi “smarter”, e degli eventuali errori<sup>88</sup>.

## 8. Privacy e altre garanzie nel trattamento algoritmico dei dati

Come si è detto in precedenza, i dati sono tutto nel “nuovo mondo” dell’AI.

I sistemi di AI, la capacità degli algoritmi di generare risposte, “predizioni”, informazioni, dipendono dalla generazione, collezione, raccolta, analisi e uso di una gigantesca quantità di dati, che riguarderanno le nostre condizioni di salute, le nostre preferenze di consumatori, le opinioni personali e politiche espresse sui social, foto, precedenti giudiziari o di rapporti con le banche e molte altre informazioni ancora. Come si legge nel Report AIHR<sup>89</sup>, «Much of the data that fuels AI systems will either be personally identified, or rife with the possibility of being re-identified using an algorithm in the event that was anonymized».

Le implicazioni per la privacy sono evidenti<sup>90</sup>, non solo come difesa della sfera di intimità e di riservatezza di ciascun soggetto, ma prestando grande attenzione all’esigenza di evitare o diminuire il rischio di un uso discriminatorio di queste informazioni, ad esempio sul piano del diritto al lavoro, dei contratti assicurativi, della stessa procedura penale.

Il recente Regolamento europeo relativo alla protezione delle persone fisiche con riguardo al trattamento dei dati personali (GDPR), sebbene non direttamente (cioè con riferimento specifico ai sistemi di AI), comincia a porre qualche paletto importante.

Si pensi all’art. 22, che afferma il principio secondo cui «L’interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici che lo riguardano o che incida in modo analogo significativamente sulla sua persona».

Sempre il GDPR, al considerando (71), nel confermare il principio prima ricordato, prevede che, qualora trattamenti automatizzati siano previsti e autorizzati, «tale trattamento dovrebbe essere subordinato a garanzie adeguate, che dovrebbero comprendere la specifica informazione all’interessato e il diritto di ottenere l’intervento umano, di esprimere la propria opinione, di ottenere una spiegazione della decisione conseguita dopo tale valutazione e di contestare la decisione»<sup>91</sup>.

<sup>88</sup> Cfr. P. ZELLINI, *La dittatura*, cit., 136.

<sup>89</sup> AIHR, cit., 18.

<sup>90</sup> Scrive A. BARICCO, *The Game*, cit., 239, «la violazione della privacy pare sistematica, e sembra essere il prezzo da pagare per i servizi che quei player (sta parlando di Google, Facebook, Amazon, Apple) ci mettono a disposizione gratuitamente. A quanto pare la regola è questa: quando è gratis, quello che stanno veramente vendendo sei tu».

<sup>91</sup> Per A. SIMONCINI, cit., 15, «ottenere una spiegazione esprime un *quid pluris* rispetto al diritto ad avere informazioni rilevanti sulla logica della decisione».

Inoltre, in tema di profilazione, «è opportuno che il titolare del trattamento utilizzi procedure matematiche o statistiche appropriate per la profilazione, metta in atto misure tecniche e organizzative adeguate al fine di garantire, in particolare, che siano rettificati i fattori che comportano inesattezze dei dati e sia minimizzato il rischio di errori e al fine di garantire la sicurezza dei dati personali secondo una modalità che tenga conto dei potenziali rischi esistenti per gli interessi e i diritti dell’interessato e che impedisca tra l’altro effetti discriminatori nei confronti di persone fisiche sulla base della razza o dell’origine etnica, delle opinioni politiche, della religione o delle convinzioni personali, dell’appartenenza sindacale, dello status genetico, dello stato di salute o dell’orientamento sessuale, ovvero che comportano misure aventi tali effetti».

Si tratta di ipotesi problematiche e diritti su cui ora insiste anche la Toronto Declaration intitolata *Protecting the right to equality and non-discrimination in machine learning systems*, documento promosso tra gli altri da Amnesty International, e aperto alla firma dal 16 maggio 2018.

Nel Preambolo viene apertamente riconosciuto il carattere ambiguo di queste nuove tecnologie, «the potential for machine learning and related systems to be used to promote human rights, but are increasingly concerned about the capability of such systems to facilitate intentional or inadvertent discrimination against certain individuals or groups of people. [...] In a world of machine learning systems, who will bear accountability for harming human rights?».

Vengono segnalate in questo senso alcune opportune metodiche di controllo e di monitoraggio che dovrebbero essere regolate e adottate sul piano normativo. In particolare, si legge nella Declaration: «i) Subjecting systems to live, regular tests and audits; interrogating markers of success for bias and self-fulfilling feedback loops; and ensuring holistic independent reviews of systems in the context of human rights harms in a live environment; ii) Disclosing known limitations of the system in question – for example, noting measures of confidence, known failure scenarios and appropriate limitations of use; iii) Ensure transparency and accountability around public sector use of machine learning systems. This must include explainability and intelligibility in the use of these technologies so that the impact on affected individuals and groups can be effectively scrutinised by independent entities, responsibilities established, and actors held to account».

In questa ottica, gli Stati dovrebbero pubblicizzare i settori di utilizzo del *machine learning* nella sfera pubblica; informare in modo chiaro e accessibile in che modo si perviene a realizzare processi decisionali basati su meccanismi “automatizzati” e di *machine learning*; documentare le azioni e le misure adottate per identificare e correggere (o almeno ridimensionare) gli impatti discriminatori o pregiudizievoli per i diritti (*rights-harming impacts*); evitare l’uso di cd. *black box systems* che non sono assoggettabili ad adeguati standards di trasparenza e *accountability*, soprattutto in contesti ad alto rischio per diritti e interessi fondamentali delle persone<sup>92</sup>.

L’impatto trasformativo (e forse rivoluzionario) sul piano sociale dello sviluppo di AI systems richiede infine che la società nella sua interezza sia preparata e coinvolta, attraverso una corretta e costante informazione, nella progressiva implementazione di questi strumenti tecnologici nell’attività istituzionale e nella vita quotidiana<sup>93</sup>.

<sup>92</sup> Sottolinea che il GDPR incoraggia e promuove una sorta di “right of an explanation” con riferimento ai sistemi algoritmici, AIHR cit., 54 ss.

<sup>93</sup> R. ANGELINI, *Intelligenza artificiale*, cit., 23.

## 9. AI, algoritmi e “società dell’anticipazione”. Problemi del processo democratico

Le potenzialità del mondo algoritmico si riversano – e non sempre ne siamo perfettamente consapevoli – sulle nostre preferenze e opzioni comportamentali, le anticipano, o forse sarebbe meglio dire che le dettano o le costruiscono<sup>94</sup>, combinando machine learning, big data, scienza comportamentale, tecniche di *nudging*<sup>95</sup>; come scrive D. Cardon<sup>96</sup>, «gli algoritmi nascono da un desiderio di autonomia e libertà. Tuttavia contribuiscono anche ad assoggettare l’internauta a quella strada calcolata, efficace, automatica, che si adatta ai nostri desideri regolandosi, in segreto, sul traffico altrui».

Quanto siamo liberi di scegliere cosa comprare, ma anche (e qui diventa un delicato problema di democrazia) cosa votare o come pensare? Qual è il confine tra ciò che chiediamo al software e ciò che ci viene indicato di chiedere?<sup>97</sup>.

Sul piano costituzionalistico in particolare, non può non attirare la nostra attenzione il rischio di *microtargeting* degli elettori.

Può influenzare, e in che termini le nostre opinioni politiche? Può incidere in qualche modo sulla nostra decisione di andare a votare oppure no? Può collocarci più o meno inconsapevolmente in gruppi di condivisione di idee e di posizioni politiche (le “*filter bubbles*” e le “*echo chambers*” di cui parla, nel suo interessante contributo, M. Fasan)?<sup>98</sup>

Quale impatto può avere tutto questo sulla democrazia elettorale dove a volte poche centinaia di voti possono determinare, tanto per fare un esempio reale non molto lontano nel tempo, l’elezione del Presidente americano? Il caso Cambridge Analytica è lì a dirci che queste preoccupazioni sono qualcosa in più di una semplice fantasia<sup>99</sup>.

<sup>94</sup> C. ACCOTO, *Il mondo dato*, cit., 86-87, parla di una “società dell’anticipazione”, di una vera e propria «*vocazione precognitiva delle tecnologie del XXI secolo, una spinta al tempo anticipato che è operazionalizzato da sensori e nuove tecnologie di processamento dei dati raccolti grazie all’intelligenza artificiale*» (cfr. anche 95-96).

<sup>95</sup> Cfr. lo Statement on Artificial Intelligence, Robotics and ‘Autonomous’ Systems, cit., 8.

<sup>96</sup> In *Che cosa sognano gli algoritmi?*, Milano, 2016, 90. Vedi anche L. ALEXANDRE, *La guerra delle intelligenze*, cit., 21, secondo cui «*i consumatori (chi non lo è) sono gli utili idioti dell’IA. Alimentiamo la macchina digitale di domani, senza esserne consapevoli*».

<sup>97</sup> Secondo M. HILDEBRANDT, *Smart technologies and the End(s) of Law. Novel Entanglements of Law and Technology*, Cheltenham, 2015, 11, «*ambienti intelligenti anticipatori cominciano a trasformare il nostro rapporto con gli artefatti. A un certo punto, diventeremo consapevoli del fatto che siamo guardati e anticipati dalle macchine e che dovremo capire come l’infrastruttura ci legge e con chi scambia questa conoscenza di noi e delle nostre preferenze (e dei rischi che questo incorpora)*».

<sup>98</sup> Cfr. C. O’NEIL, *Armi di distruzione matematica*, cit., 267 ss., secondo cui il moderno consumer marketing offre ai politici nuove strade per raggiungere elettori specifici a cui raccontare quello che vogliono sentirsi dire. Alla fine, «*queste tecnologie possono posizionare ciascuno di noi nella nostra accogliente nicchia politica*».

<sup>99</sup> In un’intervista del 4 marzo scorso, Giovanni Buttarelli, dal 2014 al vertice dell’European Data Protection Supervisor, l’authority europea che si occupa della protezione dei dati personali dei cittadini dell’Unione, ha evidenziato la sussistenza di rischi di profilazione degli utenti a scopi elettorali anche in vista del prossimo voto per le Europee. «*Tutto quello che facciamo online lascia delle tracce – ha spiegato Buttarelli – queste informazioni sono raccolte da soggetti commerciali che poi le vanno a rivendere ai partiti politici per lanciare messaggi mirati a fasce di elettorato*».

Più in generale, il rischio è quello di un impoverimento del dibattito pubblico e democratico<sup>100</sup>: nessuno dialoga più con gli altri, se non con quelli che condividono le proprie posizioni<sup>101</sup>. Il pluralismo delle opinioni non scompare, ma tuttavia viene costretto in una dimensione reciprocamente chiusa, rigida, mentre il vero pluralismo ha bisogno di mobilità delle opinioni, di disponibilità ad ascoltare le ragioni degli altri, e a farsene eventualmente convincere.

Più che di libertà di manifestazione del pensiero, in discussione è la stessa libertà di pensiero, e il suo presupposto, vale a dire una corretta e consapevole informazione.

L’esperienza del politico virtuale (SAM) in Nuova Zelanda, al quale è dedicato il saggio di F. Pacini, è in realtà una sorta di *warning* sui rischi della democrazia rappresentativa di fronte all’idea di una politica che diventa –come scrive acutamente questo A.- “pura amministrazione del consenso”, un consenso registrato quotidianamente secondo logiche di prevalenza, con modalità che sembrano idonee ad assecondare un’evoluzione tecnologica del populismo.

La saldatura tra questi due processi (disintermediazione e radicalizzazione del confronto delle opinioni) può diventare un mix letale per la tenuta dei sistemi democratici, che rischia di trasfigurare la partecipazione dei cittadini alla vita politica, economica, sociale del Paese, che è lo scopo fondamentale del progetto costituzionale di eguaglianza sostanziale (art. 3, secondo comma, Cost.), in uno scontro tra ‘blocchi’ di dati che i governanti di turno possono utilizzare più o meno opportunisticamente.

Fortunatamente non ci sono solo questi scenari distopici, e francamente inquietanti (almeno per chi ha conosciuto la democrazia costruita sulla fatica della rappresentanza e della discussione pubblica, pur con tutte le sue imperfezioni e distorsioni).

Va detto che i sistemi di AI potrebbero certamente costituire uno strumento valido per valutare (anche preventivamente) il livello di fattibilità delle leggi e degli atti normativi, migliorando così la qualità delle decisioni. Guardando poi più avanti nel tempo (quanto, è difficile prevedere), in alcuni settori le modalità di funzionamento dei sistemi di AI potrebbero configurarsi tout court come “norme auto-applicative”, capaci di prevenire i comportamenti *contra legem*, semplicemente evitando o bloccando preventivamente la loro commissione<sup>102</sup>, o adattando i margini di azione al contesto. J. Susskind, in *Future Politics*, fa l’esempio dei limiti di velocità per chi guida, che potrebbero essere implementati direttamente dal software della vettura, magari tenendo conto delle circostanze ambientali, persino degli eventuali precedenti “infortunistici” del conducente, e così via.

<sup>100</sup> Di «*fragmentation of the public sphere, ... that favour only certain types of news outlets, thereby enhancing levels of polarisation in society which can seriously jeopardise social cohesion*», parla il report *Algorithms and Human Rights*, cit., 17.

<sup>101</sup> A. BARICCO, *The Game*, 213 ss., parla di un ‘individualismo di massa’, aggiungendo questa riflessione: «*L’individualismo è sempre, per definizione, una postura contro: è sedimento di una ribellione, ha la pretesa di generare un’anomalia, rifiuta di camminare nel gregge e cammina in solitudine in controsenso. Ma quando milioni di persone si mettono a camminare in controsenso, qual è il senso giusto della strada?*».

<sup>102</sup> J. SUSSKIND, *Future Politics. Living together in a World transformed by Tech*, Oxford University Press, 2018.

## 10. L'interrogativo "intrattabile". Ci sarà una soggettività per gli AI systems o agents?

Concludo queste mie brevi riflessioni su AI e diritto con la domanda più 'estrema', quasi indicibile, e che tuttavia non si può eludere, a prescindere dal grado di attualità e di maturità tecnologica dei suoi presupposti<sup>103</sup>.

Chi sono questi AI systems o agents? Ed è chiaro che già l'uso del "chi" e non del "cosa", è una provocazione<sup>104</sup>.

È possibile immaginare una personalità legale per questi "sistemi agenti", man mano che acquisiranno e dimostreranno sempre maggiore autonomia (pur nei termini prima delineati) nello svolgere le loro funzioni intellettive e operative, e nel prendere scelte che possono avere una dimensione morale (si pensi ancora al dilemma delle collisioni inevitabili, prima richiamato)<sup>105</sup>? Quale regime giuridico possiamo assegnare alle opere dell'ingegno prodotte da una AI?<sup>106</sup>

Anche nella Risoluzione del Parlamento Europeo sulla robotica, si usa questa categoria concettuale ("sistemi agenti"), che è una sorta di via di mezzo tra "soggetti" e "oggetti"<sup>107</sup>. E l'esigenza, in quel documento, non è solo teorica o speculativa (sebbene si tratti di un documento di *soft law*): il problema è come distribuire la responsabilità civile per i danni causati dai robots.

<sup>103</sup> Correttamente, B. BISOL, A. CARNEVALE, F. LUCIVERO, *Diritti umani, valori e nuove tecnologie. Il caso dell'etica della robotica in Europa*, in *Metodo. International Studies in Phenomenology and Philosophy*, vol. 2, n. 1/2014, 250, affermano che «da un punto di vista scientifico ed epistemologico la riflessione sociale non può essere limitata all'esistente. Avremmo eliminato buona parte della realtà. Reale, infatti, non è solo ciò che si può fare ma è anche ciò che è possibile. Il fattibile e il possibile hanno infatti due campi semantici che non sempre coincidono. Relegare la riflessione etica al realismo dei fatti è, di fatto, limitante. [...] La criticità etica dei robot, perciò, riguarda anche ciò che essi rappresentano per l'essere umano in termini di immaginazione e di possibilità».

<sup>104</sup> Cfr. scrive B. ROMANO, nella Prefazione a A. PUNZI, *L'ordine giuridico delle macchine*, cit., X, secondo cui se la libertà è la «parola iniziante, capace dello scegliersi nell'originalità del futuro», con riferimento alla macchina manca proprio il "chi" di questa parola iniziante, e le operazioni della macchina sono «l'esecuzione asoggettiva di un programma senza il chi. [...] non si riesce a descrivere un giudizio giuridico che riguardi una macchina o un vivente non umano, estranei all'esercizio della libertà e quindi ad ogni presentarsi di elementi del fenomeno diritto».

<sup>105</sup> C. ACCOTO, *Il mondo dato*, cit., 76 ss., ricorda che al MIT stanno lavorando proprio su una piattaforma chiamata "the Moral Machine", con riferimento specifico alle macchine a guida autonoma, aggiungendo che «stiamo andando verso una intelligenza artificiale che è chiamata ad interagire in modi tradizionalmente riservati alla socialità umana (che diventa socialità artificiale come le macchine che non solo auto-imparano, ma che, per esempio, insegnano anche ad altre macchine). Nel fare questo, l'AI [...] sta ponendo così le basi per una profonda ridefinizione proprio del nostro concetto umano di esperienza».

<sup>106</sup> Il problema è avanzato da J. BALKIN, *The Path*, cit., 58, secondo cui «courts might treat AI-produced art as a 'work for hire' in order to minimize changes to existing copyright law». Intanto, l'estate scorsa la stampa di tutto il mondo ha dato la notizia che la famosa casa d'asta Christie's ha battuto un quadro 'generato/creato' da un sistema algoritmico, alimentato da migliaia di ritratti analoghi al 'soggetto' che era stato assegnato, vale a dire un uomo del XVII secolo. Il 'Ritratto di Edmond Belamy' è stato battuto per 432.000 dollari.

<sup>107</sup> Secondo A. CELOTTO, *I robot possono avere diritti?*, in questa Rivista, «la Risoluzione comprende che stiamo andando verso un modello in cui sarà sempre più difficoltoso legare le attività delle macchine a quella di un singolo responsabile umano. Per cui va individuata una imputabilità autonoma».

La sensazione è effettivamente che questo incedere incalzante di una “macchinalità” che simula le manifestazioni dell’umanità, come l’intelligenza, «tend to blur the line between person and instrument»<sup>108</sup>.

Questa è sicuramente una sfida enorme per il diritto, e per le sue definizioni teoriche. L’orizzonte problematico che si apre appare davvero sconfinato, e forse incontenibile all’interno di concetti che comunque – per quanto evoluti e rinnovati – pur sempre partono o si riconducono al registro dell’esperienza umana.

Non a caso vengono proposte nuove parole, che non sembrano un semplice adattamento del linguaggio al *novum* digitale: si parla di soggettività ‘elementale’, come «nuova dimensione di presenza e di azione introdotta dalle reti sensoriali, digitali e artificiali e dalle loro dinamiche operanti a livelli e scale che sono al contempo, sopra e sotto la dimensione dell’umano», o “data-driven” cioè guidate e potenziate dai dati; di “agentività”, come termine più adatto a considerare «entità che agiscono nel mondo, ma senza che abbiano di necessità le caratteristiche che, di norma, assegniamo all’azione intenzionale umana»<sup>109</sup>; di “nuove attorialità”, cioè «dei soggetti sintetici e del loro peso relazionale, del modo cioè in cui essi modificano il contesto ambientale in cui l’uomo è calato»<sup>110</sup>.

La soggettività giuridica, in fondo, è (almeno in parte) una invenzione/estensione del pensiero umano<sup>111</sup>, una delle forme di manifestazione della capacità adattiva del diritto alle esigenze pratiche poste dalle dinamiche sociali, scientifiche, culturali. L’uso di “finzioni” per aggirare le barriere della soggettività naturalisticamente intesa, è un’operazione tutt’altro che rara, e comunque in molti casi praticata e accettata, nella dimensione giuridico-normativa.

D’altronde, il diritto può scegliere di trattare qualcosa in un modo o in un altro anche per semplici ragioni di opportunità, anche se le tessere del mosaico teorico di riferimento non combaciano perfettamente.

Mi piace ricordare che uno dei primi a porre questo problema in un bellissimo saggio del 1992 è stato proprio Lawrence Solum, che è tra gli autori di questo focus tematico (con il contributo *Artificially Intelligent Law*)<sup>112</sup>. La domanda iniziale del saggio era proprio «Could an artificial intelligence become a legal person?». Una domanda (in quel momento, cioè 27 anni fa) solo teorica, avvertiva l’A., fondata

<sup>108</sup> Così R. CALO, *Robotics in American Law*, cit., 107, secondo cui «a robot is an artifact, but one holding special fascination and significance», aggiungendo altresì che «we will only understand the true impact of robots in American Law and legal institutions in the fullness of time».

<sup>109</sup> Vedi, anche per le espressioni riportate tra virgolette, C. ACCOTO, *Il mondo dato*, cit., 93 ss., secondo cui «in una prospettiva elementale, la soggettività deve essere ripensata e riconcettualizzata in quanto non più individuabile come prerogativa privilegiata di attori umani singoli. [...] Dobbiamo anche abbandonare le facili dicotomie tra umano e non umano e cominciare a pensare a un umano come una delle possibili declinazioni o concretizzazioni della soggettività elementale».

<sup>110</sup> Vedi A. VENANZONI, *Intersezioni costituzionali*, cit., 17 e 20, che parla di «innovazioni tecnologiche che non si presentano più come meri oggetti strumentali di cui servirsi ma come elementi soggettivizzati capaci di porre in essere proprie volizioni o comunque azioni destinati a riverberarsi relazionalmente sugli individui e sui gruppi sociali».

<sup>111</sup> Afferma D. SIVADON, *Macchine reti desiderio*, in F. BERARDI/BIFO (a cura di), *Cibernauti. Tecnologia, comunicazione, democrazia. Elementi di psiconautica*, Roma, 1994, 76, «Nulla dunque di meno naturale che la soggettività. Nulla di più costruito, di più elaborato, lavorato».

<sup>112</sup> Il saggio a cui si è fatto riferimento nel testo è L. B. SOLUM, *Legal Personhood for artificial intelligences*, in *North Carolina Law Review*, n. 4/1992, 1231 ss.

sul presupposto che la natura dell'intelligenza umana è computazionale, che la mente può essere modellata alla stregua di un programma che gira su un computer, e che le ricerche in tema di AI tentano proprio di sviluppare questi modelli.

Una premessa, come si è visto, non da tutti condivisa. Nella coscienza umana (con tutte le difficoltà e le incertezze nel definire cosa si voglia davvero intendere con questo termine e se esista un solo modo di declinare la 'coscienza'<sup>113</sup>), secondo molte ricostruzioni, giocano un ruolo fondamentale i sentimenti, le emozioni, i nostri corpi in movimento (la cd. "situadness"). Come scrive S. Hustvedt, «ragionare non consiste in un semplice calcolo logico, ma comporta delle emozioni»<sup>114</sup>.

Le riflessioni della Hustvedt riprendono e sviluppano l'elaborazione teorica di A. Damasio. Per Damasio «gli organismi viventi sono costruiti secondo algoritmi<sup>115</sup> e li usano per far funzionare le loro macchine genetiche, ma non sono essi stessi algoritmi. (...) Non sono linee di codice di un linguaggio di programmazione ma materia palpabile<sup>116</sup>». L'importanza di un corpo, come si è detto, di un substrato fisico-biologico che sostiene il processo fondamentale di regolazione della vita, secondo Damasio: vale a dire l'omeostasi. Omeostasi – sentimenti – valori – sistemi morali: questa è la sequenza che definisce la coscienza e l'intelligenza umana, la sua irriducibile unicità<sup>117</sup>. Diversamente, secondo Damasio, «l'intelligenza di cui danno prova quei programmi è spettacolare, ma non ha molto in comune con i processi mentali umani»<sup>118</sup>.

Nondimeno, il lavoro continua nei laboratori di ricerca proprio sulla simulazione delle reazioni emotive. Non possiamo escludere che in futuro avremo robots sempre più capaci di replicare o di offrire qualche prova di emotività. Probabilmente non saranno mai la realtà, ma solo un suo modello sem-

<sup>113</sup> Sulla difficoltà di definire la coscienza, e di costruire su questo concetto un argomento per escludere la soggettività/personalità dei sistemi di AI, v. L. B. SOLUM, *Legal Personhood*, cit., 1264-1266, il quale, rimanendo dentro la sua premessa teorica sulla struttura computazionale della mente, afferma: «*If consciousness is a computation property of the brain, then in principle we ought to be able to reproduce it with the right sort of comuter. Putting it another way, we can get consciousness out of neurons. Why not transistors?*».

<sup>114</sup> S. HUSTVEDT, *Le illusioni della certezza*, cit., 168. La stessa A. sostiene che «*i cervelli sono anche nei corpi, che a loro volta sono nel mondo. (...) Noi animali ingeriamo il mondo in molti modi, quando mangiamo, mastichiamo, respiriamo. Assorbiamo il mondo con occhi, orecchie e naso, lo assaggiamo con la lingua e ne sperimentiamo la consistenza sulla pelle. (...) Questi processi natutali sono separabili dai nostri pensieri e dalle nostre parole? (...) Non aveva forse ragione Turing quando diceva che questo modello (cioè iol modello puramente computazionale) escludeva il sesso, il cibo e lo sport e molto di quello che conta nelle nostre vite? E questa non si può forse considerare una lacuna del modello?*». Al contrario, almeno allo stato delle conoscenze attuali, «la macchina non fa alcuna esperienza».

<sup>115</sup> E fin qui sembrerebbe esserci un'assonanza tra questa riflessione di Damasio e un passaggio molto famoso di R. DAWKINS, *L'orologio cieco. Creazione o evoluzione?*, trad. it., Milano, 2016, 160, secondo cui «*se si vuol comprendere la vita, non si pensi a gel e fanghi vibranti, palpitanti, bensì alle tecnologie dell'informazione*».

<sup>116</sup> Anche secondo G. M. EDELMAN, *Sulla materia della mente*, cit., 293, «il cervello non è un calcolatore e ... il mondo non è così univocamente specificato da poter agire come un insieme di istruzioni».

<sup>117</sup> A. DAMASIO, *Lo strano ordine delle cose*, Milano, 2018, 233, secondo cui «*i valori che le nostre culture celebrano sotto forma di opere d'arte, credenze religiose, sistemi di giustizia e buon governo sono stati plasmati sulla base dei sentimenti. Se eliminassimo l'attuale substrato chimico della sofferenza e del suo contrario, il piacere e il benessere, elimineremmo il fondamento naturale dei nostri sistemi morali*».

<sup>118</sup> O.u.c., 237. Aggiunge DAMASIO che «*essi (vale a dire i sistemi di AI) sono puramente cognitivi e del tutto privi di affetti; questo significa che le loro menti intelligenti non passano attraverso le tappe intellettuali che caratterizzano la mente umana, non conoscono l'interazione con i sentimenti rievocati, con i sentimenti che accompagnano il nostro presente e con i sentimenti legati alla nostra visione del futuro*».



pre più perfezionato e “ravvicinato” all’originale<sup>119</sup>. Ma sarà davvero questo un discrimine ancora netto e valido, dal punto di vista del diritto e della considerazione sociale, di fronte allo sviluppo delle condizioni di interattività tra esseri umani e agenti artificiali, e dalla capacità di fare cose intelligenti di questi ultimi? In altre parole, il fatto che l’intelligenza artificiale non sarà mai uguale (in termini “qualitativi”) a quella umana<sup>120</sup>, che l’autonomia operativa di tali sistemi resterà priva di una percezione mentale (*id est*, di una “esperienza”) dei valori morali implicati in determinate scelte, costituirà sempre una differenza chiara ed insuperabile sul profilo della qualificazione giuridica di questi “oggetti-agenti”, come se la soggettività non possa essere un concetto differenziato e plurale?

Torniamo al lavoro di Solum. Era molto problematico, denso di riferimenti all’uso di altre *legal fictions* o di zone grigie con riferimento alla personalità giuridica. È vero che il diritto ha spesso “inventato” maschere della soggettività, allargando progressivamente, rispetto alla persona fisica capace di intendere e di volere, l’area del rilevante moralmente (e giuridicamente): pensiamo solo al concepito, agli animali, agli alberi nel celebre lavoro di Christopher Stone<sup>121</sup>, più in generale (guardando ad oggi) potremmo aggiungere il clima, le cose della natura (recentemente un fiume in Nuova Zelanda). Non solo quindi figure che alla fine sono proiezioni (o “anticipazioni”) della persona naturale e della vita umana, come le società, le generazioni future, il concepito, forse gli stessi animali in quanto esseri senzienti; ma anche cose, oggetti della natura, risorse dell’ecosistema, legate alla natura umana essenzialmente da un rapporto di necessaria contestualità, nel senso che rappresentano pezzi dello scenario in cui scorre la vita degli esseri umani.

Voglio citare alcuni passaggi della conclusione di quello che Solum stesso chiamò un ‘esperimento di pensiero’ che ci avrebbe portato oltre le nostre intuizioni e i nostri giudizi consolidati: «An answer to the question whether artificial intelligences should be granted some form of legal personhood cannot be given until our form of life gives the question urgency. But when our daily encounters with artificial intelligence do raise the question of personhood, they may change our perspective about how the question is to be answered. [...] in a modern, pluralist society, the disagreement about ultimate questions is profound and persistent. Resolution of hard cases in the political and judicial spheres requires the use of public reason. We have no realistic alternative but to seek principled compromise based on our shared heritage of toleration and respect».

Da questa eredità provengono obiezioni effettivamente forti alla possibilità di riconoscere una personalità legale agli AI systems. Solum stesso ammette che questa prospettiva «must be located in the indefinite future» (ma, appunto, era ancora il 1992)<sup>122</sup>; ciò nonostante, non è un buon motivo per

<sup>119</sup> S. HUSTVEDT, o.u.c., 182.

<sup>120</sup> Cfr., sul punto, ancora G. M. EDELMAN, *Sulla materia della mente*, cit., 346-351, secondo il quale «*Che i calcolatori effettuino operazioni logiche è indiscutibile; il guaio è che la logica da sola non rappresenta il pensiero più di quanto gli eventi fisici del sommare due numeri su un pallottoliere assomiglino a ciò che succede nel cervello di un matematico quando esegue dei calcoli o quando crea. (...)*». Per K. KELLY, *The Myth of A Superhuman AI*, in <https://blockchannel.com/>, 25 aprile 2017, l’intelligenza ha tante possibili manifestazioni; «*Tra cent’anni, il termine intelligenza sarà come il termine neve per un eschimese. Avremo altri cento modi di descriverla per distinguere le sue varietà*».

<sup>121</sup> C. D. STONE, *Does Trees Have Standing? Law, Morality, and the Environment* (1972), ripubblicato da Oxford University Press, 2010.

<sup>122</sup> Analogamente, parla di “potential long-term concerns”, il Report DEG, cit., 12-13.

non parlarne e per non confrontarsi con l'asprezza e la complessità (teorica e pratica) dei problemi che pone<sup>123</sup>.

Solum riassume così le tre obiezioni: 1. *Als are not humans*<sup>124</sup>; 2. *The Missing-Something Argument*, e cioè i sistemi di AI non hanno caratteristiche come coscienza, intenzionalità, emozioni, interessi propri, la cd. "aboutness"; 3. *Als ought to be property*, sono artefatti, il prodotto del lavoro umano<sup>125</sup>.

Come si può vedere, esse sono il precipitato di una lunga tradizione giuridica, e contengono molto di vero e di intuitivo, si riferiscono a concezioni radicate sul piano sociale e culturale. Tuttavia, sono argomenti che rischiano di reggersi l'uno sull'altro, di apparire quasi 'autoreferenziali', e ad ogni modo 'provvisori', cioè legati al contesto attuale delle possibilità tecnologiche in tema di AI.

Ma tra 20, 30, 50 anni, o forse prima? L'argomento della mancanza di coscienza o di sentimenti sarà davvero sempre inattingibile man mano che le ricerche e gli sviluppi sugli AI agents o systems riusciranno a raggiungere livelli più elevati di 'replicazione' dei comportamenti (e degli atteggiamenti) umani? La prima obiezione poi non ha impedito di pensare e di riconoscere normativamente una qualche forma di personalità legale a "entities" sicuramente non umane (come cose della natura, clima, animali) o 'non ancora' umane (generazioni future, soprattutto se le guardiamo alla luce del famoso paradosso della 'non identità' di Derek Parfit<sup>126</sup>).

Infine, l'argomento della proprietà è il meno 'stabile' di tutti, o comunque appare in diversi modi aggirabile. Scrive Solum: «First, slaves can be emancipated. If we concede that Als come into the world as property, it does not mean that they must remain so. Second, even slaves (ovviamente è un ragionamento paradossale, perchè tale condizione almeno oggi è in se, e radicalmente, anti-costituzionale) can have constitutional rights, be those rights ever so poor as compared to the rights of free persons»<sup>127</sup>. In altre parole, si potrebbe pensare di riconoscere ad un AI anche un set minimale di contenuti della libertà e dell'eguaglianza; la personalità, in una situazione limite come quella evocata, potrebbe essere 'graduata', differenziata, secondo logiche diverse da quella "tutto o niente". Peraltro, la questione (qui solo accennata in via introduttiva, e che certo merita ben altro approfondimento) ha un suo risvolto, o meglio un punto rovesciato di osservazione.

Il modo giuridico di considerare robots e sistemi di AI sarà inevitabilmente condizionato dalle molteplici interazioni che questi 'artefatti' tecnologici avranno con gli esseri umani in vari settori, alcuni dei quali con forti implicazioni relazionali ed emotive (pensiamo al campo della salute, dell'assistenza alle persone vulnerabili, come anziani e disabili, alla possibilità, pare ancora lontana nel tempo, di sexbots).

Ryan Calo parla di «social valence», e Jack Balkin di «substitution effect». In altri termini, incontrare (e interagire con) robots e AI agents sul proprio cammino della vita può portare a proiettare sugli

<sup>123</sup> Invece, per G. M. EDELMAN, *Sulla materia della mente*, cit., 304, la risposta all'interrogativo se «è possibile costruire un artefatto dotato di coscienza?» è: «*si, in linea di principio, ma la soluzione dei problemi pratici che la creazione di una coscienza di ordine superiore comporta è talmente al di là da venire che è inutile preoccuparsene per il momento*» (va detto che anche Edelman scrive nel 1992).

<sup>124</sup> Per L. B. SOLUM, o.c., 1259, «this argument suggests that our very concept of person is inextricably linked to our experience of a human life».

<sup>125</sup> O.u.c., 1258 ss. Seguendo questa obiezione, rileva Solum, «*Als should never be more than the property of their makers. Put differently, the objection is that artificial intelligence, even if persons, are natural slaves*».

<sup>126</sup> D. PARFIT, *Ragioni e persone*, trad. it., Milano, 1984, 462 ss., 447-448, 482.

<sup>127</sup> O.u.c., 1279.

stessi emozioni, desideri, responsabilità, obiettivi in modo analogo a quanto avviene ad esempio per gli animali; a maggior ragione se questi ‘artefatti’ saranno sempre più in grado di mimare in modo efficace le condizioni sensomotorie che costituiscono l’elemento essenziale dell’esperienza umana. Questo “effetto”, secondo Balkin, «may be incomplete, contextual, unstable, and, above all, opportunistic. People may treat robot as a person (or animal) for some purposes and as an object for others»<sup>128</sup>. In ogni caso però, questo fenomeno può senz’altro avere qualche riflesso sulla sistemazione giuridica di tali “tools/subjects”, sulla loro considerazione/qualificazione in termini culturali e sociali<sup>129</sup>.

Mi rendo conto che non ho dato risposte. Ma sono convinto che non sia ancora il momento di facili conclusioni sui diritti delle macchine (addirittura il diritto di voto per Ray Kurzweil) o –al contrario– sulla impossibilità logica (*ora e sempre*) di pensare ad una soggettività ‘artificiale’. Adesso è il momento di porsi degli interrogativi, di formulare ipotesi, di cominciare a testare su queste ipotesi le categorie razionali e argomentative che il diritto ci ha messo (o può metterci) a disposizione.

Ci aspetta un lungo cammino: come dice R. Calo, «Robotics today is in the midst of a sea change, advancing at a breakneck pace. The field is headed toward a Cambrian explosion, referring to the great speed of the field’s evolution and diversification»<sup>130</sup>. Il riferimento al Cambriano è suggestivo ed evocativo; è, in sé, un abbozzo di riflessione, il segno di una consapevolezza. Un ‘mistero’, lo definì lo stesso Darwin, una combinazione ‘inaspettata’ di progresso e contingenza, qualcosa di imprevedibile, almeno a priori, che tuttavia si è rivelato fondamentale per la comparsa di Homo sapiens sulla scena della biodiversità<sup>131</sup>.

Paragonare allora lo stato attuale delle ricerche e delle applicazioni in tema di AI ai “mari del Cambriano” è una metafora al tempo stesso affascinante e inquietante; significa stiamo percorrendo sentieri davvero incogniti, e che non tutto può essere oggi previsto e definito oggi. Sarebbe un peccato di presunzione, una semplificazione inutile.

## 11. Prime (e provvisorie...) conclusioni. Il diritto di fronte ad una sfida radicale e non temporanea

Insomma, per concludere, il compito che attende il diritto di fronte agli scenari inediti (talvolta finanche “impensabili”) delineati dallo sviluppo delle tecniche e dei sistemi di AI, appare effettivamente e

<sup>128</sup> J. BALKIN, *The Path of Robotics Law*, cit., 56-58. Per un ragionamento simile, v. L. B. SOLUM, *Legal Personhood*, cit., 1286. Con riguardo a tale preoccupazione vi è chi propone che «l’ampiezza dell’utilizzo dell’IA in una determinata professione dovrebbe essere inversamente proporzionale al valore del coefficiente di umanità che lo svolgimento di quella professione richiede: quanto più alto sia questo valore, tanto più l’IA dovrà avere un ruolo meramente strumentale» (così R. Angelini, *Intelligenza artificiale*, cit., 21).

<sup>129</sup> Secondo L. B. SOLUM, *Legal Personhood*, cit., 1260, se immaginiamo un mondo in cui «we interact frequently with AIs that possess human qualities, but lack any semblance of human biology (...), our concept of a person may change in a way that creates a cleavage between human and person». Non a caso, il progetto Robolaw prevedeva una serie di vincoli tra cui il divieto di costruire robot con fattezze indistinguibili dagli esseri umani, e il divieto di utilizzare i robot umanoidi in chiave sostitutiva delle persone nei legami affettivi: v. C. SALAZAR, *Umano, troppo umano*, cit., 262.

<sup>130</sup> R. CALO, *Robotics*, cit., 108.

<sup>131</sup> Per queste considerazioni, v. lo splendido libro di T. PIEVANI, *La vita inaspettata*, Milano, 2011, 51 ss.

radicalmente “trasformativo”<sup>132</sup>. Stiamo parlando di un processo che mette in gioco domande fondamentali sul futuro dell’umanità, sul “tipo di mondo (che si) desidera creare ed abitare”.

Nel farci queste domande, che non possono essere semplicemente evitate o negate (non avrebbe molto senso), dobbiamo essere coscienti che il tema ha una portata “intertemporale”, sia nel senso che l’AI non sarà una fase passeggera nell’evoluzione tecnologica, ma un modo irreversibile di ridefinire le forme della nostra esistenza; sia nel senso che quello che decidiamo o consentiamo oggi avrà un impatto anche sulle generazioni che verranno dopo di noi, *ora* e *domani*.

Simoncini propone nel suo contributo un modello di “precauzione costituzionale”. In altri contesti, ho insistito anch’io sulla necessità del diritto di assumere un’impronta intergenerazionale come connotato di razionalità delle sue manifestazioni. Non tanto un “diritto della paura” (secondo la celebre espressione di Cass Sunstein<sup>133</sup>), quanto un diritto che sente il dovere di pensare al futuro, di renderlo possibile, di mantenere nel tempo le condizioni essenziali della convivenza umana; un diritto che ha colto l’enorme forza trasformativa dell’agire umano, e che “responsabilmente” e “ragionevolmente” si preoccupa e si fa carico delle conseguenze nel tempo delle decisioni di oggi e dei problemi che è chiamato a regolare<sup>134</sup>.

L’AI è sicuramente uno dei temi che appartengono alla *intergenerational issue*<sup>135</sup>.

In questo senso, nel rapporto tra AI e diritto quello del diritto non è uno sguardo neutro, “da nessun luogo”<sup>136</sup>, a meno di non volersi arrendere alla visione pessimistica e nichilista di un diritto come vuota forma, “senza destinazione: che va e va, ma non sa ‘perché e ‘verso dove’ muova”<sup>137</sup>: ci sono principi, consolidazioni, esperienze, a cominciare da quelli che si riannodano al linguaggio costituzionale, e ai suoi contenuti fondamentali, come pure (per noi) ai valori rilanciati dalle strutture normative europee (pensiamo soprattutto alla Carta dei diritti fondamentali dell’UE<sup>138</sup>).

Certo, il tema è globale e non può che essere affrontato in una dimensione che superi i confini del diritto statale (ma anche di un diritto di dimensioni sovranazionali ma limitate come può essere quello europeo).

Forse, la prima sfida per il diritto (e per il diritto costituzionale in particolare) è proprio questa, ed è il tratto dominante di quasi tutte le nuove emergenti *issues* del nostro tempo, come ad esempio le implicazioni del *climate change*, i dilemmi della genetica.

<sup>132</sup> Cfr. ancora L. ALEXANDRE, *La guerra delle intelligenze*, cit., 289: «la neurorivoluzione che si sta sviluppando davanti ai nostri occhi va ben oltre il problema del lavoro. Non è nemmeno una questione economica, né prima di tutto una questione politica. Mette in discussione gli esseri umani in quelle che sono le loro caratteristiche più intime e preziose».

<sup>133</sup> C. SUNSTEIN, *Il diritto della paura*, trad. it., Bologna, 2010.

<sup>134</sup> Sia consentito il rinvio a A. D’ALOIA, *Generazioni future (diritto costituzionale)*, in *Enc. Dir., Annali*, vol. IX, Milano 2016, 332 ss., part. 336, 356 ss.

<sup>135</sup> Cfr. S. PRATESI, *Generazioni future? Una sfida per i diritti umani*, Torino, 2007, 29 ss., 59 ss.; e T. SERRA, *L’uomo programmato*, Torino, 2003, 36, secondo la quale «occorre cominciare a preoccuparsi non solo della domanda ‘quale mondo per l’uomo di domani?’, ma anche dell’altra, ancora più complessa, ‘quale uomo per il mondo di domani?’».

<sup>136</sup> L’espressione riportata nel testo è di C. SALAZAR, *Umano, troppo umano*, cit., 266/267.

<sup>137</sup> Vedi N. IRTI, *Nichilismo giuridico*, Bari-Roma, 2004, 8.

<sup>138</sup> Per B. BISOL, A. CARNEVALE, F. LUCIVERO, *Diritti umani, valori e nuove tecnologie*, cit., 235, nel dibattito europeo su scienza e tecnologia, la Carta dei diritti fondamentali dell’UE può rappresentare un punto di riferimento molto significativo.

In un articolo del 1960, intitolato *Man-Computer Symbiosis*<sup>139</sup>, uno dei padri dell’informatica, Joseph Licklider, nell’immaginare quelli che sarebbero stati i progressi delle macchine alla fine non riusciva a dare scadenze temporali sicure per questi progressi (parlava di 15, 20, 500 anni); aggiungeva però che questi anni, gli anni dello sviluppo dell’IA, sarebbero stati i più creativi ed eccitanti nella storia dell’umanità.

Su questa consapevolezza, che porta con sé anche un certo brivido come tutto ciò che non è interamente prevedibile, io vorrei chiudere richiamando una riflessione di Pietro Piovani, con la quale A. Punzi chiude il suo lavoro sull’ordine giuridico delle macchine<sup>140</sup>.

In un saggio del 1967, *Salus a Machina*<sup>141</sup>, esprimeva un auspicio; che cioè, dal *timor machinae*, possa svilupparsi un *initium sapientiae*, un nuovo itinerario civile della razionalità. Così concludeva Piovani: «L’involontaria pedagogia della macchina esorta alla radicalizzazione elementare del principio della vita: alla fine (letteralmente alla fine) può essere la macchina a rimettere in moto, con la sua semplificazione radicale, il congegno della coscienza. La sua potenzialità distruttiva è forse la più valida potenza sotterrica per l’umanità novecentesca. Se, grazie alla macchina scatenata, l’uomo imparerà di nuovo a tremare (nel senso, come a me piace intendere, di ragionare, ripensare...), sarà salvo».

---

<sup>139</sup> In *IRE Transactions on Human Factors in Electronics*, vol. HFE-1, march 1960.

<sup>140</sup> *Op. cit.*, 405.

<sup>141</sup> P. PIOVANI, *Salus a machina*, in *Ethica. Rassegna di filosofia morale*, 1967, 1, 44-45.